ESC Cardiovascular Research (2026) **122**, 1589–1608
https://doi.org/10.1093/cvr/cvag140

Proteomic signatures as biomarkers of atherosclerosis burden

Lanyue Zhang [1], Murad Omarov [1], LingLing Xu[1], Barnali Das[2], Hong Luo[3], Stefanie M. Hauck[4], Agnese Petrera[4], Zhi Yu[5,6], Sascha N. Goonewardena[7,8], Eleftheria Zeggini[9,10], Annette Peters[3,11,12,13], Martin Dichgans [1,13,14,15], Venkatesh L. Murthy[7], Barbara Thorand[3,11,12], and Marios K. Georgakis [1,6,14]*

[1]Institute for Stroke and Dementia Research (ISD), LMU University Hospital, LMU Medizin, LMU Munich, Feodor-Lynen-Str. 17, Munich 81377, Germany; [2]Urological Diseases Research Center, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA; [3]Institute of Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health (GmbH), Ingolstädter Landstr. 1, Neuherberg 85764, Germany; [4]Metabolomics and Proteomics Core, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Ingolstädter Landstr. 1, Neuherberg 85764, Germany; [5]Clinical and Translational Epidemiology Unit, Department of Medicine, Massachusetts General Hospital, 100 Cambridge Street, Boston, MA 02114, USA; [6]Program in Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA; [7]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA; [8]Division of Cardiovascular Medicine, VA Ann Arbor Health System, 2215 Fuller Road, Ann Arbor, MI 48105, USA; [9]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, Neuherberg 85764, Germany; [10]Technical University of Munich (TUM), TUM University Hospital, TUM School of Medicine and Health, Ismaninger Str. 22, Munich 81675, Germany; [11]Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine, Ludwig-Maximilians-Universität, Marchioninistr. 15, Munich 81377, Germany; [12]German Center for Diabetes Research (DZD), Partner München-Neuherberg, Ingolstädter Landstr. 1, Neuherberg 85764, Germany; [13]German Centre for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Pettenkoferstr. 8a/9, Munich 80336, Germany; [14]Munich Cluster for Systems Neurology (SyNergy), Feodor-Lynen-Str. 17, Munich 81377, Germany; and [15]German Center for Neurodegenerative Diseases (DZNE), Feodor-Lynen-Str. 17, Munich 81377, Germany

Received 26 September 2025; revised 22 February 2026; accepted 3 April 2026; online publish-ahead-of-print 29 June 2026

Time for primary review: 55 days

Aims

Atherosclerosis is currently evaluated by imaging, but scalable circulating biomarkers to detect its presence and quantify overall burden are lacking. We sought to define plasma proteomic signatures that reflect the systemic burden of atherosclerosis.

Methods and results

In UK Biobank, we trained machine-learning models on plasma proteomics in a nested, propensity score–matched case–control sample (1666 cases; 1666 controls; Olink Explore 3072, 2920 proteins) to derive four AtheroBurden signatures: one based on the entire proteome (all 2920 proteins) and three biologically informed subsets—genetically anchored (n = 402), atherogenesis-related (n = 680), and artery-enriched (n = 248). We then computed individual-level signatures in 41 200 disease-free participants and validated their performance through (i) correlation with carotid plaque burden via ultrasound (n = 1712); (ii) prospective major adverse cardiovascular events (MACEs) prediction over 13.7 year follow-up using multivariable Cox models and incremental prediction over Systematic COronary Risk Evaluation version 2 (SCORE2) [ΔC-index; net reclassification improvement (NRI)], as well as in the external Cooperative Health Research in the Region of Augsburg (KORA) S4 (n = 1361) and Age1 (n = 796) cohorts; and (iii) longitudinal trajectories analysis in 1210 participants with 3 serial proteomic measurements. All four AtheroBurden signatures robustly discriminated prevalent atherosclerotic disease (receiver operating characteristic area under curve up to 0.91, 95% CI: 0.89–0.93) with their levels significantly increasing with the number of clinically affected vascular beds. We found significant correlations with carotid ultrasound–measured plaque burden. In prospective analysis, each signature exhibited strong associations with incident MACE [hazard ratio (HR) per standard deviation increase in whole-proteome signature: 1.75, 95% CI: 1.68–1.83; HR for Q4 vs. Q1: 3.18, 95% CI: 2.8–3.61], providing significant improvements in risk discrimination (ΔC-index: +0.044; P < 0.0001) and reclassification [NRI: 0.12 (95% CI: 0.09–0.15) at a 10% risk threshold] beyond SCORE2. These results were replicated in the KORA S4 and KORA-Age1 cohorts, where the signatures were associated with risk of future myocardial infarction and stroke. Furthermore, longitudinal analyses demonstrated steeper annual increases in proteomic signature among individuals with higher burden of vascular risk factors and individuals who experienced MACE over follow-up.

* Corresponding author. Tel: +49 89 4400 46127; fax: +49 89 4400 46040, E-mail: marios.georgakis@med.uni-muenchen.de

Conclusion Plasma proteomic signatures effectively capture atherosclerotic burden and improve cardiovascular risk prediction in asymptomatic individuals. They may complement existing risk stratification by serving as a scalable and accessible blood-based screening tool to identify individuals more likely to have subclinical atherosclerosis and may benefit from confirmatory imaging and earlier prevention strategies.

Keywords Proteomics • Atherosclerosis • Machine learning • Risk prediction • Cardiovascular disease

1. Introduction

Cardiovascular disease (CVD) remains the leading global cause of death and disability,[1,2] driven primarily by atherosclerosis,[3] a progressive, lipid-driven inflammatory process that silently accumulates over decades before culminating in clinical events such as myocardial infarction (MI) and stroke.[4,5] Despite advances in prevention and treatment, many individuals with a high atherosclerosis burden remain undiagnosed until the occurrence of a major cardiovascular event, underscoring a critical gap in early detection and prevention.[6] The continued rise in the incidence of cardiovascular events[7,8] further emphasizes the need to refine current paradigms of risk assessment.

Current prevention strategies rely on population-level algorithms, such as Systematic COronary Risk Evaluation version 2 (SCORE2) and the pooled cohort equations, which estimate cardiovascular risk based on demographic and clinical variables including age, sex, blood pressure, and cholesterol levels.[9–11] While widely adopted, these models do not directly quantify atherosclerotic burden and offer limited resolution in individual risk assessment. Direct detection of subclinical atherosclerosis remains dependent on imaging modalities including angiography, computed tomography, magnetic resonance imaging, positron emission tomography, or ultrasound.[12–14] While informative, these techniques are constrained by procedural risks, radiation exposure, availability, and the need for specialized personnel.[15] Circulating biomarkers may help address some of these limitations by serving as scalable screening tools to identify individuals more likely to have subclinical atherosclerosis, prioritize them for imaging-based confirmation, and support cardiovascular risk stratification and longitudinal monitoring. However, existing circulating biomarkers, such as C-reactive protein[16] or cardiac troponins,[17] primarily reflect systemic inflammation or myocardial injury and fall short of directly assessing atherosclerotic plaque burden or progression.

Circulating proteins may serve as real-time indicators of pathophysiological processes.[18] Recent advances in proteomic technologies enable the simultaneous quantification of thousands of proteins,[19,20] providing a window into dynamic, tissue-specific pathophysiological processes. Integrating these data through machine learning has uncovered proteomic signatures predictive of early stages of neurodegenerative disease,[21–25] cancer,[26–28] diabetes,[29–31] autoimmune disease,[32,33] and mortality risk through proteomic ageing clocks.[34] While previous studies have shown the potential for plasma proteomics in improving prediction of specific cardiovascular outcomes,[35–41] these efforts have largely been optimized for event prediction and were not designed to directly quantify underlying systemic atherosclerosis burden or its progression. In parallel, proteomics studies targeting subclinical atherosclerosis have often been conducted in cohorts of limited size[42–44] and have focused on specific vascular territories or imaging phenotypes,[45–48] thereby limiting the systematic assessment of overall atherosclerotic burden and its longitudinal evolution.

Here, we leveraged plasma proteomics from the UK Biobank (UKB) and two independent cohorts to develop and validate four biologically informed proteomic signatures of atherosclerotic burden (AtheroBurden). Using machine learning, we constructed four signatures based on data from 1666 cases with established atherosclerotic disease and 1666 age- and sex-matched controls: WholeProteome (derived from the entire proteome), Genetic [genetically anchored proteins identified via Mendelian randomization (MR)], Mechanistic (proteins implicated in atherogenesis), and Arterial (artery-enriched proteins). We analysed associations between the signatures and carotid plaque burden measured by imaging. Subsequently, we evaluated the ability of these signatures to predict incident cardiovascular events in 41 200 disease-free UKB participants (median follow-up 13.7 years), and further validated externally in Cooperative Health Research in the Region of Augsburg (KORA) S4 (n = 1,361, median follow-up 15.1 years) and KORA-Age1 (n = 796, median follow-up 6.8 years). Finally, we assessed the longitudinal trajectories of these four signatures across three serial time points spanning a median of 12.5 years and investigated how signature trajectories are influenced by baseline cardiovascular risk factors and the occurrence of future cardiovascular events.

2. Methods

2.1 Study populations and datasets

The UKB represents a prospective cohort study encompassing 502 421 individuals from the general UK population.[49] Between March 2006 and October 2010, participants aged 37–73 years attended one of 22 assessment centres across Scotland, England, and Wales.[49,50] Each participant completed a touchscreen questionnaire, had physical measurements taken, and provided blood, urine, and saliva samples at baseline. Detailed information about the UKB protocol can be found at http://www.ukbiobank.ac.uk. Baseline plasma proteomics at the UKB baseline visit (Instance 0) were available for 54 219 participants; after excluding participants with >30% missing values across all measured proteins, 44 788 participants remained and constituted the UKB analytic proteomics cohort, from which the datasets described below were derived.

2.1.1 Discovery dataset (case–control study)

The discovery dataset, used to develop the AtheroBurden scoring system, was drawn from the UKB analytic proteomics cohort. We identified 1666 cases with established atherosclerotic disease (detailed in the 'Atherosclerosis Ascertainment' section) and matched controls without atherosclerotic events at baseline or during follow-up. For control selection, each case was first matched to six candidates based on age and sex using propensity score matching (R package MatchIt, v 4.5.5);[51,52] from each set of 6 matched controls, we then selected the control with the fewest International Classification of Diseases, 10th Revision (ICD-10) diagnosis codes, thereby reducing potential confounding due to comorbidity burden and improving case–control comparability, yielding the final 1:1 matched case–control dataset of 3332 participants.

2.1.2 Internal validation dataset (prospective cohort study)

To assess the performance of the AtheroBurden scores developed from the discovery dataset, we analysed data from UKB

participants with proteomics data and no history of atherosclerotic CVD that were not included in the discovery dataset. We included individuals aged 40–70 years, resulting in 41 200 participants. Assuming that a score capturing atherosclerosis burden in asymptomatic individuals should be associated with future risk of major adverse cardiovascular events (MACEs), we assessed associations of AtheroBurden scores with the occurrence of incident MACE over a median follow-up of 13.7 years.

2.1.3 Longitudinal analysis dataset in UKB

To evaluate temporal changes in AtheroBurden scores and their associations with cardiovascular risk profiles, we utilized repeated proteomic measurements available for a subset of UKB participants. A total of 1210 individuals, derived from the COVID-19 repeat-imaging study, had at least 2 proteomic measurements, allowing for longitudinal analysis.

2.1.4 Carotid plaque subset in UKB

To assess associations of the developed AtheroBurden scores with subclinical atherosclerosis burden, we used data from a follow-up imaging visit of UKB participants initiated in 2014. A subset of 82 340 participants underwent carotid ultrasound as part of a comprehensive assessment. From this group, 19 499 individuals provided a total of 177 757 carotid ultrasound images, which were subsequently processed for plaque evaluation. Ultrasound imaging was performed using a standardized protocol to capture bilateral carotid arteries.[53] A total of 1712 participants had both carotid ultrasound and plasma proteomics data, enabling the investigation of associations between AtheroBurden scores and carotid plaque presence and burden.

2.1.5 KORA cohorts

To externally validate our findings, we utilized data from KORA studies,[54] specifically the KORA S4 and KORA-Age1 cohorts.[31,39] Ethical approval was granted by the local ethics committee, and all participants provided written informed consent. KORA S4 was conducted between 1999 and 2001, enrolling 4261 participants, of which 1361 participants aged 55–74 years were included in our analysis due to the availability of proteomics data and follow-up data. KORA-Age1 recruited 1079 participants aged 65–93 years, with proteomics and follow-up data available for 796 individuals. Both cohorts were used to investigate the relationship between proteomics and cardiovascular outcomes, specifically incident MI and stroke. Among the 1361 participants in KORA S4, 207 were also part of the KORA-Age1 cohort, though these participants were assessed at different time points. In sensitivity analyses, we excluded the overlapping participants from both cohorts to ensure independence between datasets.

2.2 Proteomic profiling and preprocessing

2.2.1 Plasma proteomics in UKB

Blood samples were primarily collected from UKB participants during their baseline visit (Instance 0), with additional samples gathered from members of the UKB Pharma Proteome Consortium and individuals in the COVID-19 repeat-imaging study. Proteomic profiling was performed using the antibody-based Olink® Proteomics proximity extension assay (PEA) technology in ethylenediaminetetraacetic acid (EDTA)–anticoagulated plasma samples.[55]

At baseline (Instance 0), proteomic profiling was performed using the Olink® Explore 3072 platform, which encompasses eight distinct panels: Cardiometabolic, Cardiometabolic II, Inflammation, Inflammation II, Neurology, Neurology II, Oncology, and Oncology II. This comprehensive platform enabled quantification of 2923 unique proteins across 54 219 participants.[56] Samples were representative of the broader UKB population, with 93% of European ancestry. Protein levels were provided as normalized protein expression (NPX) values, generated by log-transforming counts normalized to extension controls.[57] Assessments indicated that protein expression levels were minimally affected by protein batch, study centre, and genetic principal components. Detailed protocols for sample handling, processing, and quality control (QC) are available online.[56] For follow-up assessments (Instances 2/3), the Olink® Explore 1536 platform was employed, resulting in measurements of 1463 proteins. Despite differences in panel coverage between baseline and follow-up assessments, the fundamental profiling technology and QC procedures remained consistent, ensuring methodological comparability across time points.[57]

After excluding three proteins [glioma pathogenesis-related protein 1 from the Oncology II panel, Procollagen C-endopeptidase enhancer 1 (PCOLCE) from the Cardiometabolic panel, and nucleophosmin (NPM1) from the Neurology panel] that were missing in more than 30% of participants in the final study cohort, the remaining missing values were imputed using a normal distribution method as previously described.[58] The mean of this imputation distribution was adjusted by subtracting 1.8 SD from the mean of the abundance distribution of all proteins in one sample. The standard deviation of the imputation distribution was set to 0.3 times the standard deviation of the abundance distribution.

2.2.2 Plasma proteomics in KORA

Proteomics data for both KORA S4 and KORA-Age1 cohorts were also measured in EDTA plasma using Olink® Proteomics PEA technology, covering 276 protein biomarkers from CVD (CVD-II and CVD-III) and inflammation panels. The data processing, including QC and normalization, was performed by the KORA team as previously described.[31,39] Proteins with more than 25% of values below the limit of detection (LOD) or with missingness were excluded. For proteins present in multiple panels, the version with fewer values below LOD and a lower inter-assay coefficient of variation was retained. After applying these QC criteria, 233 unique proteins passed QC in KORA S4, while 243 proteins passed QC in KORA-Age1. Due to differences in the specific protein panels used between the KORA cohorts and UKB, 232 proteins from KORA S4 and 242 proteins from KORA-Age1 overlapped with those measured in UKB; proteomic signatures were therefore recalculated using these overlapping proteins for external validation in KORA.

2.3 Outcomes definition

2.3.1 Atherosclerosis ascertainment

Clinical diagnoses and surgical records were utilized as proxies to identify individuals with presence of atherosclerotic disease. Atherosclerosis was ascertained by identifying events across multiple vascular beds, including coronary, extra- and intracranial, aortic, peripheral, and other arterial sites. Curated disease phenotypes were defined using clinical diagnosis codes from the International Classification of Diseases, 9th and 10th revisions (ICD-9 and ICD-10), as well as surgical procedure codes from the Office of Population Censuses and Surveys, 4th revision. Diagnosis dates were obtained from linked individual participant data. Incident events due to atherosclerotic disease were ascertained from hospital inpatient data summaries (fields 41270, 41271, 41272) as outlined in Supplementary material online, *Table S1*. Prevalent events were defined as those occurring before the participant's baseline visit when a blood sample was collected. Individuals with corresponding prevalent events for each outcome were considered as cases. Individuals without any experienced atherosclerotic events at baseline and during follow-up were considered as controls and subsequently underwent propensity score matching to construct the discovery dataset. For each individual, atherosclerotic events were evaluated across the five vascular beds described above.

The presence of an event in any vascular bed scored 1 point, resulting in atherosclerotic burden levels ranging from 0 (no events) to 2 (events in two or more vascular beds).

2.3.2 MACE outcome definitions

The outcome in the internal validation cohort included incident traditional three-point MACE, which comprised non-fatal acute MI (AMI), non-fatal ischaemic stroke (IS), and cardiovascular death (CV death). ICD-9 and ICD-10 codes for each endpoint are listed in Supplementary material online, *Table S2*, ascertained from linked Hospital Episode Statistics and death registries. For each participant, follow-up began at their baseline visit to the UKB assessment centre, where clinical information and blood samples were collected. The first occurrence of a MACE event was recorded as the primary endpoint for the composite outcome analysis, ensuring each participant contributed only once. For participants without a MACE event, follow-up was censored at the earliest of non-cardiovascular death, or the last available hospital inpatient record (31 October 2022 for England, 31 August 2022 for Scotland, and 31 May 2022 for Wales). Mortality data were available until 31 October 2022, and participants without events were censored on these respective dates. When analysing individual components of MACE (AMI, IS, and CV death) as separate outcomes, we included each participant's first occurrence of each specific event type.

In the KORA cohorts, the endpoint was the first validated MI or IS. MIs were ascertained via the Augsburg MI Registry: events before 31 December 2000 followed World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease adjudication, and those from 1 January 2001 used European Society of Cardiology/American College of Cardiology criteria. MIs outside the registry's area or age limits (>74 years, >84 years from 2009) were identified through follow-up questionnaires and confirmed with hospital or physician records; fatal MI cases were identified through death certificates or autopsy reports. Non-fatal ISs were initially identified through self-reports and validated with medical records, while fatal strokes were identified through death certificates or autopsy reports. KORA S4 participants had their baseline examination in 1999–2001, first follow-up examination in 2006–08 and second follow-up examination in 2013–14. Furthermore, postal questionnaires were sent out in 2008–09 and 2016; KORA-Age1 participants had their baseline visit in 2008–09, first follow-up visit in 2012, and a postal questionnaire was sent to them in 2016.

2.3.3 Carotid plaque assessment

Carotid ultrasound images were analysed using the deep learning model described by Omarov *et al.*,[59] which assesses carotid plaque presence and the number of plaques in the left and right carotid arteries. Plaques were defined as focal protrusions into the arterial lumen with a thickness greater than 50% of the surrounding carotid intima-media thickness,[60] with plaque presence determined by the identification of at least one plaque in either carotid artery. Plaque burden was assessed based on the total number of plaques detected in both carotid arteries and categorized as 0 for participants with no plaques, 1 for those with a single plaque, and 2 for those with two or more plaques.

2.4 Covariates and SCORE2 calculation

2.4.1 Demographic and covariates

Baseline variables used in our analyses included age, sex, ethnic background, smoking and drinking status, systolic blood pressure (SBP), diastolic blood pressure (DBP), cholesterol levels, body mass index (BMI), kidney function [estimated glomerular filtration rate (eGFR)], glycated haemoglobin A1c (HbA1c), first-degree family history of CVD, and history of diabetes and hypertension. Detailed definitions of these variables, including UKB field IDs, are provided in Supplementary material online, *Table S3*. Self-reported information on ethnic background, smoking status (current, previous, or never), drinking status (current, previous, or never), and first-degree family history of CVD (heart disease and/or stroke in any parent or sibling) was obtained from baseline questionnaires. Blood pressure was measured during the baseline visit, and the average of two readings was used. BMI was calculated as weight in kilograms divided by height in metres squared (kg/m^2). Cholesterol levels, including total cholesterol (TC), low LDL cholesterol (LDL-C), HDL cholesterol (HDL-C), and triglycerides, were measured from fasting blood samples. Kidney function was assessed using eGFR calculated from serum creatinine and Cystatin C levels using the Chronic Kidney Disease Epidemiology Collaboration equation (2021).[61] History of diabetes and hypertension was assessed based on medication use and hospital records. Information on the use of glucose-lowering medications, antihypertensive medications, and lipid-lowering medications was obtained from participant medication data. Hospital records were reviewed to identify prior diagnoses using relevant ICD-9 and ICD-10 codes, with specific codes provided in Supplementary material online, *Table S4*. For the KORA cohorts, similar demographic and clinical variables were collected and defined as previously described.[62,63]

2.4.2 SCORE2 calculation

We estimated the 10 year risk of MACE for each participant using the SCORE2 algorithm,[64] based on individual factors such as age, sex, SBP, TC, HDL-C, and smoking status. Participants aged 40–70 years without MACE were included in this analysis. The linear predictor for each participant was calculated using sex-specific regression coefficients from the SCORE2 working group.[64] To better align observed and predicted risk, we applied log hazard ratios (HRs) from the SCORE2 sensitivity analysis that specifically excluded UKB participants (as reported in Supplementary Table 8 of the SCORE2 publication).[64] For absolute risk calculation, following the approach described in previous studies,[65] these linear predictors were converted into calibrated 10 year risks using the SCORE2 recalibration formula with scaling factors for low-risk European regions (as reported in Supplementary methods Table 4 of the SCORE2 publication).[64]

2.5 Machine learning model development

To explore the potential of plasma proteomics to deliver novel biomarker signatures for atherosclerosis, we developed the AtheroBurden scoring system using machine learning (ML) classifiers in the discovery dataset. The process involved selecting relevant protein features, constructing diagnostic models using various ML algorithms, evaluating their performance, and generating continuous AtheroBurden scores from the best model.

2.5.1 Protein feature pre-selection

To leverage atherosclerosis biology while minimizing confounding from late-stage organ damage signals, we designed four protein panels:

2.5.1.1. Whole-proteome panel

Without prior feature selection, we included all 2920 plasma proteins to construct an ML model predicting the probability of atherosclerosis presence.

2.5.1.2. MR-derived protein panel

We conducted a two-sample MR analysis to identify proteins that are genetically influenced by predisposition to coronary artery disease (CAD), providing causal evidence for their potential roles in

atherosclerosis pathways, with CAD serving as the exposure and plasma protein levels as the outcome. Genetic instruments were selected from the largest available CAD Genome-Wide Association Study (GWAS) summary statistics by Aragam *et al.*,[66] filtered for genome-wide significance ($P < 5 \times 10^{-08}$), and further clumped to retain independent variants ($r^2 < 0.001$, 10,000 kb window). These instruments were then matched to the Coronary ARtery DIsease Genome-wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics (CARDIoGRAMplusC4D) 1000 Genomes–based GWAS summary statistics.[67] This dataset does not include UKB data, ensuring that there was no overlap between exposure and outcome datasets. After filtering and matching, 217 single-nucleotide polymorphisms were selected for the MR analysis, with all necessary data, including beta values, obtained from the CARDIoGRAMplusC4D.

Plasma protein data were sourced from the UKB Pharma Proteomics Project, which measured 2940 plasma proteins in 54 219 participants. GWAS summary statistics for these data are publicly available via Synapse.[57] (https://www.synapse.org/Synapse:syn51365303) The MR analysis was conducted using the R package *TwoSampleMR (v 0.5.6)*, employing the random effect inverse variance weighted method for estimating causal effects. Using these data, the MR analysis identified 402 proteins ($P < 0.05$) whose levels were genetically influenced by predisposition to CAD, suggesting their potential role as causal mediators in the disease pathway.

2.5.1.3. Atherosclerosis-related protein panel

To identify proteins specifically associated with atherosclerosis, we utilized the Enrichr platform,[68] querying relevant terms and pathway databases for gene sets related to atherosclerosis. This search yielded 52 gene sets, from which we compiled a comprehensive list of genes ($n = 3312$) associated with atherosclerosis. These annotations were then mapped to the UKB Olink proteome, resulting in 680 atherosclerosis-related proteins, which constituted the atherosclerosis-related protein panel used for model development.

2.5.1.4. Artery-enriched protein panel

This panel focused on proteins with specific or elevated expression in arterial tissues, hypothesized to be closely related to atherosclerotic lesions. We obtained gene expression data from the Genotype-Tissue Expression (GTEx) project (Release V8, dbGaP Accession phs000424.v8.p2),[69] which provided comprehensive tissue-specific bulk RNA-seq expression profiles across various human tissues, including vascular tissues. We grouped three vascular tissues—coronary, aorta, and tibial artery—together as the vascular group, while all other organs were grouped as the non-vascular group. Genes were considered artery-enriched if their expression levels in the vascular group were at least three-fold higher than in the non-vascular group. We then mapped these genes to the plasma proteomics data from the UKB, resulting in an artery-enriched protein panel of 248 proteins used for model development.

2.5.2 ML classifiers

We utilized Python packages *scikit-learn (v 1.3.2)*, *catboost (v 1.2.5)*, *lightgbm (v 4.2.0)*, and *xgboost (v 2.0.3)*, to implement a range of ML techniques, including logistic regression, random forest, elastic net regression (ElasticNet), support vector machine (SVM), and multilayer perceptron (MLP). Additionally, gradient boosting classifiers such as Light Gradient Boosting Machine (LightGBM), categorical boosting (CatBoost), and eXtreme Gradient Boosting (XGBoost) were employed. These classifiers were designed to predict whether participants belonged to Class 1 (diagnosed with atherosclerotic events at baseline) or Class 0 (event-free). ML models were established using a discovery dataset created using propensity score matching based on age and sex to select healthy controls ($n = 1666$) for participants with prevalent atherosclerotic events ($n = 1666$). This matching technique helps mitigate potential nonlinear confounding effects. We then randomly split the discovery dataset into training (80%) and testing sets (20%), with stratification ensuring balanced distribution of atherosclerotic events in both sets.

All models were trained and validated using ten iterations of five-fold stratified cross-validation on the training set, with the dataset resampled for each iteration to ensure robustness. Performance was evaluated using accuracy, precision, recall, and area under the receiver operating characteristic curve (ROC-AUC), providing comprehensive insights into classification effectiveness and error patterns. Hyperparameters for the cross-validated models were optimized using Optuna,[70] an automated framework, with each configuration undergoing the same cross-validation strategy. Algorithm-specific search spaces were defined, encompassing learning rates (10^{-5} to 10^{-1}), regularization parameters (C values from 10^{-5} to 10), tree depths (3–10), number of estimators (50–200), and other model-specific parameters. Performance was assessed using average ROC-AUC and other relevant metrics, and optimal hyperparameters were selected based on configurations achieving the highest cross-validated scores. CatBoost was selected as the best-performing model based on its highest average ROC-AUC and stability (consistency of performance across testing sets). Hyperparameter specifications for all evaluated models are provided in Supplementary material online, *Table S5*. To compare the performance of the selected CatBoost model with a traditional risk prediction approach, a separate logistic regression model was developed using SCORE2 variables (age, sex, TC, HDL-C, SBP, and smoking status) in the discovery dataset, with performance assessed using ROC-AUC. Statistical significance of ROC-AUC differences was evaluated using the DeLong-test. Feature importance was assessed using SHapley Additive exPlanations (SHAP) values, which quantify each feature's contribution to the model's predictions.[71] Clinical benchmarking and sensitivity analyses were conducted to assess robustness and potential reliance on correlated biomarker pathways. For clinical benchmarking, we fit logistic regression models using SCORE2 variables alone and an extended set additionally including N-terminal pro-brain natriuretic peptide (NT-proBNP), apolipoprotein B (ApoB), and lipoprotein(a). To assess potential redundancy between biologically related natriuretic-peptide analytes, we repeated CatBoost model training after excluding either natriuretic peptide B (NPPB) or NT-proBNP. In addition, we fit a restricted-proteome CatBoost model excluding four assays related to natriuretic peptides and apolipoproteins (NPPB, NT-proBNP, APOB, and apolipoprotein(a)]. All benchmarking and sensitivity analyses used the same discovery dataset, train/test split, and evaluation framework as the primary analysis.

2.5.3 Generating continuous AtheroBurden signature

The CatBoost classifier[72] was constructed using protein expression profiles as input features, with models trained to discriminate between individuals with and without atherosclerotic disease. Following comprehensive hyperparameter optimization through five-fold cross-validation, the final classifier was applied to the entire UKB cohort to generate continuous risk predictions. The raw prediction values obtained directly from the CatBoost algorithm—representing the untransformed linear combination of weighted protein features—were subsequently standardized as *Z*-scores (centred at zero with a standard deviation of 1) to facilitate inter-individual comparability. Four AtheroBurden signatures were derived from the respective protein panels: AtheroBurden-Arterial, based on the artery-enriched panel; AtheroBurden-Mechanistic, based on the atherosclerosis-related panel; AtheroBurden-Genetic, based on the MR-derived panel; and AtheroBurden-WholeProteome, based on the whole-proteome panel. All models were trained in the discovery dataset and subsequently applied to the entire UKB cohort.

For longitudinal validation using follow-up measurements and external validation in independent cohorts, restricted versions of the AtheroBurden signatures were generated to address differential protein coverage. For UKB participants assessed at follow-up timepoints (Instances 2/3) where the Olink® Explore 1536 platform was employed, restricted signatures were computed by applying the original prediction algorithm with missing value assignments (coded as NA) for proteins not measured on the restricted platform. An identical methodological approach was implemented for external validation in the KORA cohorts. To quantify potential information loss resulting from reduced protein coverage, equivalent restricted signatures were generated in the baseline UKB cohort (Instance 0) using only proteins available across all platforms. Correlation analyses were subsequently conducted to evaluate the proportion of variance in the full signatures that could be explained by these restricted protein models.

2.6 Statistical analysis

Population characteristics were summarized as mean ± standard deviation (SD) for normally distributed variables, median [interquartile range (IQR)] for skewed variables, and *n* (%) for categorical variables. Missing clinical data were imputed using predictive mean matching via the R package *mice (v 3.16.0)*. Continuous variables were imputed using predictive mean matching, binary variables via logistic regression, and ordinal variables with a proportional odds model. Imputation was based on age and sex (no missing values) to enhance data quality and repeated five times for robustness. We evaluated pairwise associations between baseline protein NPX levels and traditional cardiovascular risk factors using Spearman rank correlations. For proprotein convertase subtilisin/kexin type 9 (PCSK9), correlations with clinical ApoB and LDL-C were additionally stratified by lipid-lowering medication status. Cox proportional hazards regression models were applied to evaluate the association between AtheroBurden scores and time to MACE among participants without baseline MACE. Three models were constructed: Model 1 adjusted for age, sex, and ethnicity background. Model 2 adjusted for age, sex, TC, HDL-C, SBP, and smoking (based on SCORE2 variables); and Model 3 further adjusted vascular risk factors (VRFs) included age, sex, ethnicity background, SBP, BMI, smoking status, drinking status, LDL-C, triglycerides, eGFR, HbA1c, diabetes, hypertension, and family history of CVD. Multiple comparisons were addressed using false discovery rate (FDR) correction to control for type I error. Additionally, Logistic and Poisson regressions were employed to assess AtheroBurden signatures' relationships with carotid plaque presence and burden, respectively. The same covariate adjustment strategy (Models 1–3) was applied in these analyses.

To assess the added value of AtheroBurden signatures over SCORE2, we evaluated discrimination improvement using concordance indices (C-index), calculated with the concordance.index function (*survcomp package*, *v 1.52.0*, R). C-index differences (ΔC-index) were compared using the cindex.comp function, reporting *P*-values and 95% CI. Time-dependent ROC (timeROC) curves were generated at 10 year follow-up points to track model performance over time. Kaplan–Meier survival curves were used to estimate cumulative MACE incidence across AtheroBurden signature quartiles, with log-rank tests performed for group comparisons. Calibration of the SCORE2 model, with and without AtheroBurden signatures, was evaluated using calibration plots comparing observed 10 year Kaplan–Meier estimates and predicted probabilities within deciles of predicted risk. Reclassification metrics included net reclassification improvement (NRI), category-free NRI (cfNRI), and integrated discrimination improvement (IDI). The NRI analysis employed two established clinical thresholds (7.5 and 10%) derived from SCORE2 risk stratification guidelines, selected based on their validated clinical utility in cardiovascular risk assessment. These thresholds were applied as population-level cut-points to evaluate the overall reclassification performance of proteomic models when added to traditional risk factors. NRI was calculated using the package *nricens (v.1.6)*[73] with confidence intervals (CIs) and *P*-values based on 1000-fold bootstrap standard errors. The cfNRI and IDI metrics were calculated using the *survIDINRI (v.1.1-2)*.[74] Sensitivity analyses examined AtheroBurden signatures' associations with individual MACE components (AMI, IS, and CV death) with FDR correction.

The longitudinal progression of AtheroBurden signatures was assessed using linear mixed-effects models with time since baseline (in years) as a continuous variable. The scores were derived from AtheroBurden scoring systems based on available proteomic data and were repeatedly measured for the same individuals at three time points: baseline (Instance 0) and two follow-ups (Instances 2 and 3). The mixed-effects models included random intercepts to account for individual-level variability, with fixed effects for standardized baseline risk factors and time.

In the external validation analysis, Cox proportional hazards models, adjusted using the same variables as previous analysis, were applied to examine the associations between restricted AtheroBurden signatures and incident MI and stroke. Kaplan–Meier survival curves were applied to estimate cumulative MI and stroke incidence across AtheroBurden scores quartiles in KORA S4, while improvements in discrimination were evaluated by changes in C-index when incorporating restricted AtheroBurden scores into SCORE2 variables across both cohorts. Sensitivity analysis excluded overlapping participants between KORA S4 and Age1 to maintain independence. Statistical power calculations for external validation analyses were conducted using R package *powerSurvEpi* (*v 0.1.3*),[75] with an alpha level of 0.05.

All statistical analyses were performed using R (version 4.3.3), and ML procedures were conducted in Python (version 3.9.10). A two-sided $P < 0.05$ was considered statistically significant.

3. Results

3.1 Summary of the study design

The study design is summarized in *Figure 1*. A detailed study workflow, including data processing, ML model development, and validation steps, is provided in Supplementary material online, *Extended Figure S1*. Of the 502 421 participants enrolled in the UKB, a total of 44 788 participants [54% female, median age 58 years (IQR: 50–64 years)] met our inclusion criteria, after excluding participants with >30% missing proteomic data (see Supplementary material online, *Extended Figure S2*). To develop proteomic signatures of atherosclerosis (AtheroBurden signatures), we leveraged four sets of proteins (see Supplementary material online, *Extended Figure S3*) and trained ML models to discriminate the 1666 cases with established atherosclerotic disease from 1666 age- and sex-matched controls (discovery dataset). The developed AtheroBurden signatures were subsequently tested for associations with incident MACE over a median follow-up of 13.7 years (n = 41 200), followed by external validation in KORA S4 (n = 1,361, median follow-up 15.1 years) and KORA-Age1 cohorts (n = 796, median follow-up 6.8 years). Baseline characteristics of participants in the development cohort (UKB) and both validation cohorts (KORA S4 and KORA-Age1) are presented in *Table 1*. We further explored associations of the derived signatures with imaging evidence of atherosclerosis on carotid ultrasound (n = 1712), as well as serial changes across three timepoints and longitudinal progression patterns stratified by both baseline SCORE2 risk categories and incident MACE status (n = 1210) in subsamples of the UKB.64

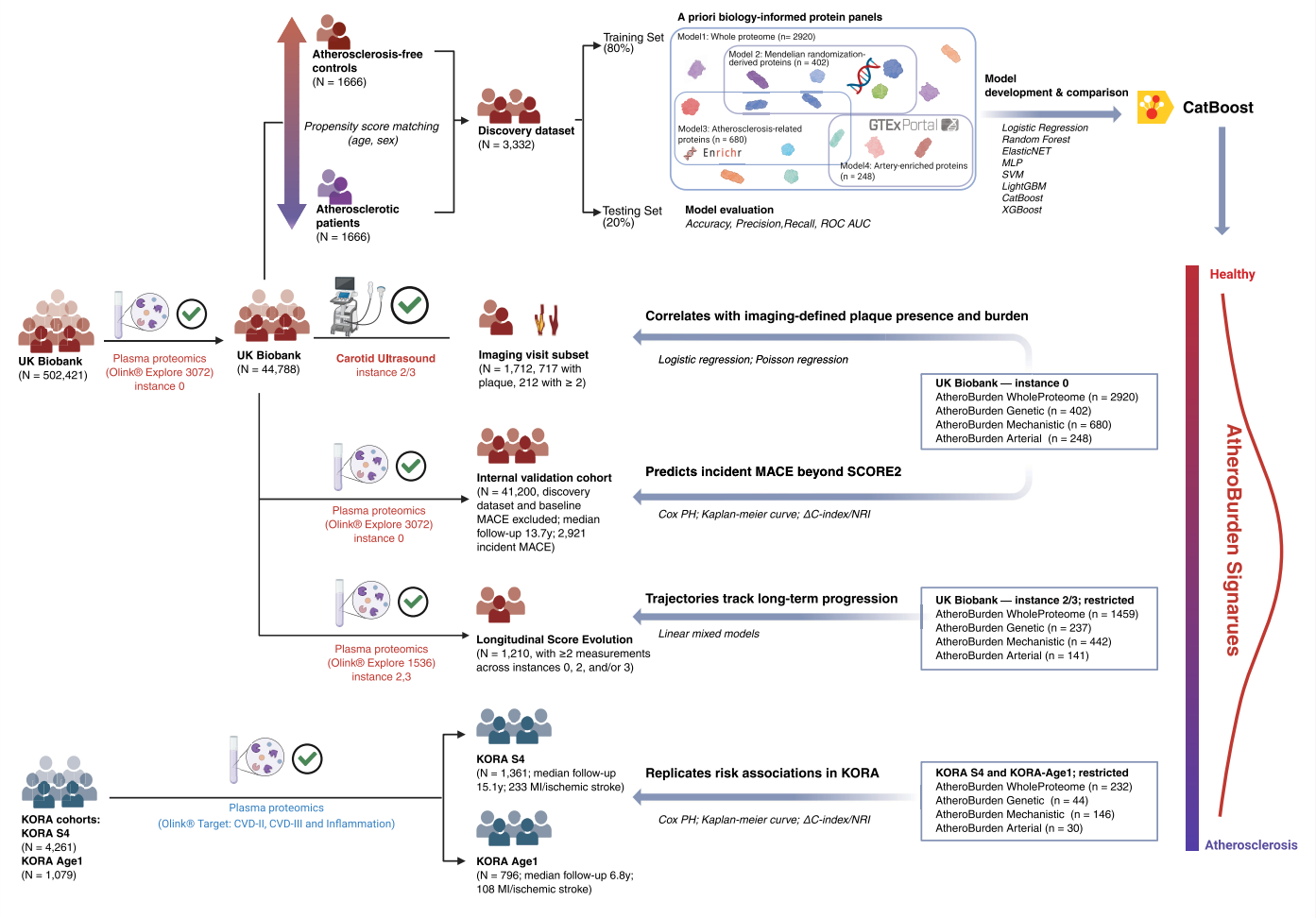

Figure 1 Overview of the study design and analytical approaches. The diagram illustrates the methodological approach implemented for protein-based atherosclerosis burden quantification. Machine learning models were trained in a UKB discovery dataset (n = 3332; 1666 atherosclerosis cases and 1666 age- and sex-matched controls) using 4 biologically informed protein panels. The resulting AtheroBurden signatures were validated in a disease-free UKB cohort (n = 41 200; median follow-up 13.7 years), assessed for association with carotid plaque burden (n = 1712), evaluated longitudinally (n = 1210), and externally validated in the KORA S4 (n = 1361) and Age1 (n = 796) cohorts.

3.2 Development of AtheroBurden proteomic signatures

To construct proteomic signatures of atherosclerosis burden, we developed ML classifiers using a case–control discovery dataset comprising 1666 participants with an established diagnosis of atherosclerotic CVD and 1:1 age- and sex-matched controls [median age 63 years (IQR: 59–66 years), 30% female, Supplementary material online, *Table S6*]. Atherosclerotic disease was defined by diagnostic codes encompassing coronary, cerebrovascular (including carotid), aortic, and peripheral arterial manifestations (see Methods). We evaluated the diagnostic performance of eight ML models—Logistic Regression, Random Forest, ElasticNET, MLP, SVM, LightGBM, CatBoost, and XGBoost—using four sets of proteins. The four protein sets were selected to represent different levels of biological relevance to atherosclerosis (see Supplementary material online, *Extended Figure S3*): (i) the whole proteome (2920 proteins); (ii) 402 proteins with evidence of causal association with genetic predisposition to CAD as derived from MR analyses (MR-derived panel); (iii) 680 proteins coded by atherosclerosis-related genes as curated from literature-based evidence according to the EnrichR platform[68] (atherosclerosis-related panel); and (iv) 248 proteins overexpressed in the aorta, coronary or tibial arteries, as detected in transcriptomic analyses across 54 tissues in GTEx[69] (artery-enriched panel). The list of proteins included in every set is provided in Supplementary material online, *Table S7*. Across ten iterations of five-fold cross-validation, CatBoost consistently outperformed the other tested models in accuracy, precision, discrimination, and recall (see Supplementary material online, *Extended Figure S4* and *Table S8*). While MLP and ElasticNET achieved higher performance than CatBoost for the artery-enriched panel in certain iterations, their results were inconsistent and exhibited significant variability. In contrast, CatBoost demonstrated robust and stable performance across all four panels, maintaining superior accuracy and reliability compared with other models (see Supplementary material online, *Extended Figure S4* and *Table S8*). CatBoost also outperformed all other models in accuracy in the testing set (see Supplementary material online, *Extended Figure S5* and *Table S9*). We therefore selected CatBoost-derived models for subsequent analyses.

As shown in *Figure 2A*, the selected CatBoost models achieved high true positive and true negative rates across all panels in the testing set. Compared with a baseline model using SCORE2 variables (ROC-AUC: 0.80), the proteomic panels significantly improved

Table 1 Baseline characteristics of participants for each the three cohorts analysed for this study

Characteristic	UKB N = 44 788	KORA S4 N = 1361	KORA-Age1 N = 796
Age at recruitment (years), median (Q1, Q3)	58.00 (50.00, 64.00)	63.00 (59.00, 68.00)	76.00 (70.00, 81.00)
Sex, *n* (%)			
Female	24 234 (54%)	685 (50%)	423 (53%)
Male	20 554 (46%)	676 (50%)	373 (47%)
Ethnicity, *n* (%)		NA	NA
Asian	982 (2%)		
Black	1075 (2%)		
White	41 692 (93%)		
Mixed	285 (1%)		
Other or Unknown	754 (2%)		
Born in Germany	NA	982 (72%)	NA
BMI (kg/m^2), median (Q1, Q3)	26.78 (24.19, 29.90)	27.95 (25.65, 30.83)	27.93 (25.51, 30.70)
DBP (mm Hg), median (Q1, Q3)	82.00 (75.00, 89.00)	80.00 (73.50, 87.50)	76.00 (69.50, 83.00)
SBP (mm Hg), median (Q1, Q3)	138.00 (126.00, 152.00)	135.00 (121.50, 148.00)	138.00 (124.50, 150.50)
TC (mmol/L), Median (Q1, Q3)	5.62 (4.89, 6.38)	6.27 (5.53, 6.97)	5.53 (4.81, 6.15)
HDL-C (mmol/L), median (Q1, Q3)	1.40 (1.18, 1.67)	1.44 (1.19, 1.75)	1.45 (1.20, 1.68)
LDL-C (mmol/L), median (Q1, Q3)	3.50 (2.94, 4.08)	3.94 (3.27, 4.60)	3.28 (2.76, 3.90)
Triglycerides (mmol/L), median (Q1, Q3)	1.49 (1.06, 2.14)	1.36 (0.99, 1.93)	1.41 (1.02, 1.99)
eGFR (mL/min/1.73 m^2), median (Q1, Q3)	95.06 (84.89, 104.57)	84.50 (74.73, 92.29)	71.70 (59.43, 83.03)
HbA1c (mmol/mol), median (Q1, Q3)	35.30 (33.00, 38.00)	37.71 (35.52, 40.98)	37.71 (35.52, 40.98)
Previous smoking, *n* (%)	15 806 (35%)	510 (38%)	307 (39%)
Current smoking, *n* (%)	4816 (11%)	187 (14%)	34 (4%)
Previous drinking, *n* (%)	1726 (4%)	NA	NA
Current drinking, *n* (%)	40 489 (90%)	993 (73%)	519 (65%)
Blood pressure medication, *n* (%)	10 384 (23%)	455 (34%)	535 (67%)
Cholesterol lowering medication, *n* (%)	8152 (18%)	134 (9.9%)	199 (25%)
Diabetes, *n* (%)	2009 (5%)	89 (7%)	112 (14%)
Hypertension, *n* (%)	11 068 (25%)	737 (54%)	599 (75%)
Family history of CVD, *n* (%)			NA
Yes	25 015 (56%)	282 (21%)	
No	18 501 (41%)	817 (60%)	
Unknown	1272 (3%)	250 (18%)	

NA, not applicable.

discrimination. The atherosclerosis-related, MR-derived, and whole-proteome panels achieved comparable enhancements (ROC-AUCs: ~0.91, DeLong-test $P < 0.001$; *Figure 2B*; Supplementary material online, *Table S10*), while the artery-enriched panel resulted in a modest, non-significant improvement (ROC-AUC: 0.84, DeLong-test $P = 0.09$; *Figure 2B*; Supplementary material online, *Table S10*). To understand the contributions of individual proteins, we calculated SHAP values across each panel. Renin (REN), NT-proBNP, NPPB, and PCSK9 consistently emerged as the top contributors to the atherosclerosis-related, MR-derived, and whole-proteome panels (*Figure 2C*). Given the biological overlap between NPPB and NT-proBNP, we performed drop-one sensitivity analyses by retraining models after excluding either analyte; discrimination remained stable (ΔROC-AUC <0.01; Supplementary material online, *Table S9*). In additional benchmarking analyses, an extended clinical comparator including SCORE2 components plus NT-proBNP, clinical ApoB and clinical lipoprotein (a) achieved an area under curve (AUC) of 0.85 (95% CI 0.82–0.87; Supplementary material online, *Table S10*), whereas a restricted-proteome model excluding natriuretic peptides and apolipoprotein-related analytes retained high discrimination (AUC 0.90, 95% CI 0.88–0.92; Supplementary material online, *Table S10*). Correlations between the top-ranked proteins and traditional risk factors are shown in Supplementary material online, *Extended Figure S6* and *Table S11*.

We subsequently applied these CatBoost models to generate four complementary signatures (AtheroBurden-WholeProteome, -Genetic, -Mechanistic, and -Arterial) for all UKB participants with available proteomic data. The density distributions of all AtheroBurden signatures showed a clear rightward shift in participants with atherosclerotic disease and a corresponding leftward shift in disease-free participants, reflecting higher and lower scores relative to the population mean, respectively (*Figure 2D*). Furthermore, the signatures captured the burden of atherosclerosis, as illustrated by higher scores among participants with evidence of atherosclerotic disease in two or more vs. one arterial bed (*Figure 2E*).

3.3 Association with plaque presence and burden

To assess associations with subclinical atherosclerosis, we examined whether baseline AtheroBurden signatures were associated with carotid plaque detected by ultrasound. As carotid imaging

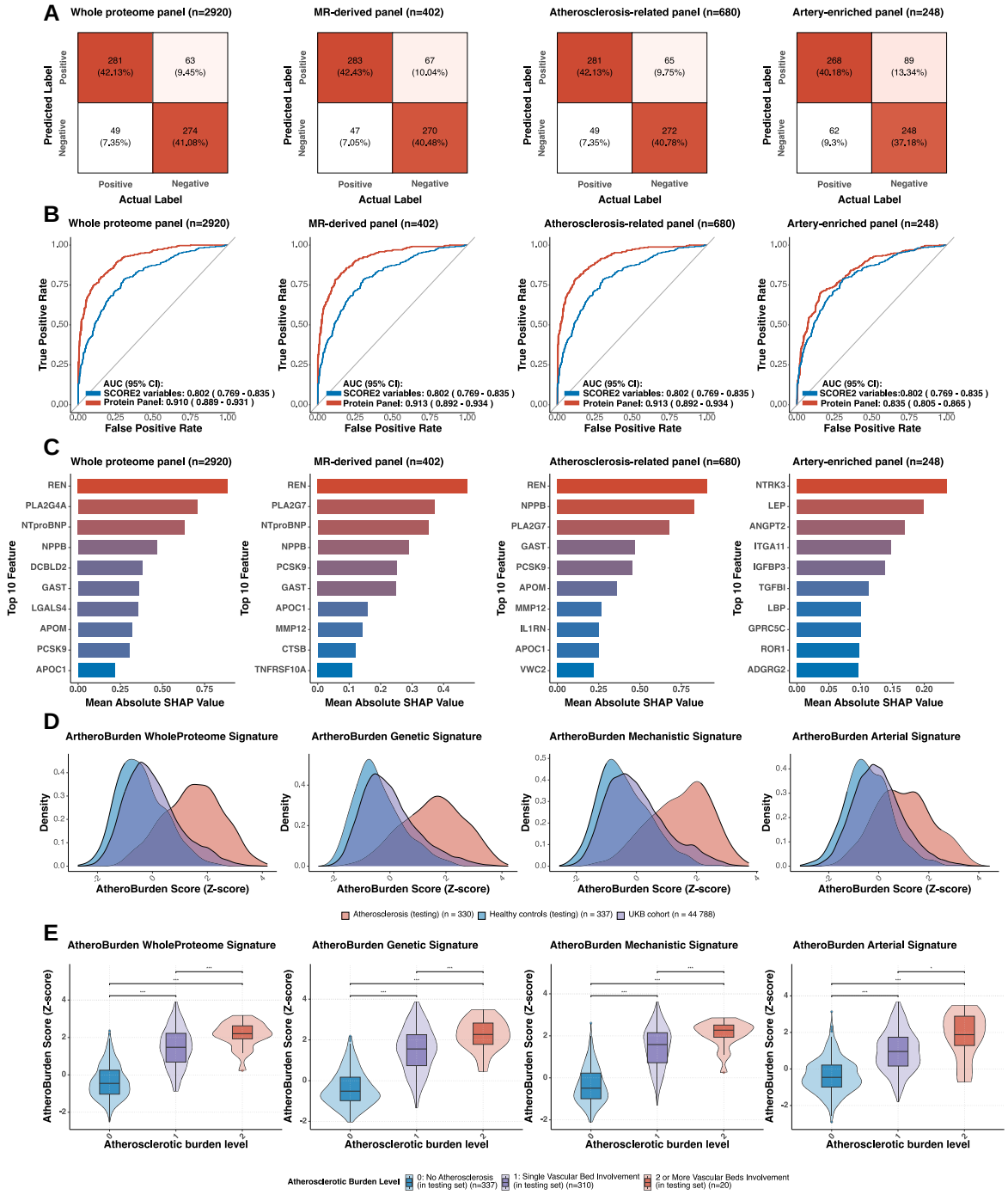

Figure 2 Evaluation of machine learning–derived proteomic signature for atherosclerosis detection and burden assessment across four protein panels. (*A*) Confusion matrices of classification performance across protein panels. The confusion matrices summarize the classification outcomes for each protein panel, illustrating proportions of true positives, true negatives, false positives, and false negatives. These results reflect the overall accuracy and error distribution of the models. (*B*) ROC curves for model evaluation. The ROC curves in the testing set illustrate the predictive performance of each protein panel. Each plot includes two curves: one representing the performance of the respective protein panel and the other showing the predictive capacity of cardiovascular risk factors included in the SCORE2 algorithm, used as a comparator. The AUC values and their 95% CI (calculated using the DeLong method) are reported for each curve. (*C*) Shapley (SHAP) values identify the top 10 contributing proteins. The bar plots display the mean absolute SHAP values for the top 10 proteins contributing to each model, ranked in descending order. (*D*) Density distributions of AtheroBurden signatures stratified by atherosclerotic status. The density plots depict the distributions of AtheroBurden signatures for healthy controls vs. atherosclerotic cases, as well as within the UKB cohort. (*E*) Violin-box plots of AtheroBurden signatures stratified by the number of affected vascular beds as a measure of atherosclerotic burden (*$P < 0.05$, ***$P < 0.001$ between indicated groups). *P*-values were obtained from two-sided Welch's *t*-tests. The middle line represents the median, boxes indicate the IQR (25th–75th percentiles), and whiskers extend to 1.5 times the IQR.

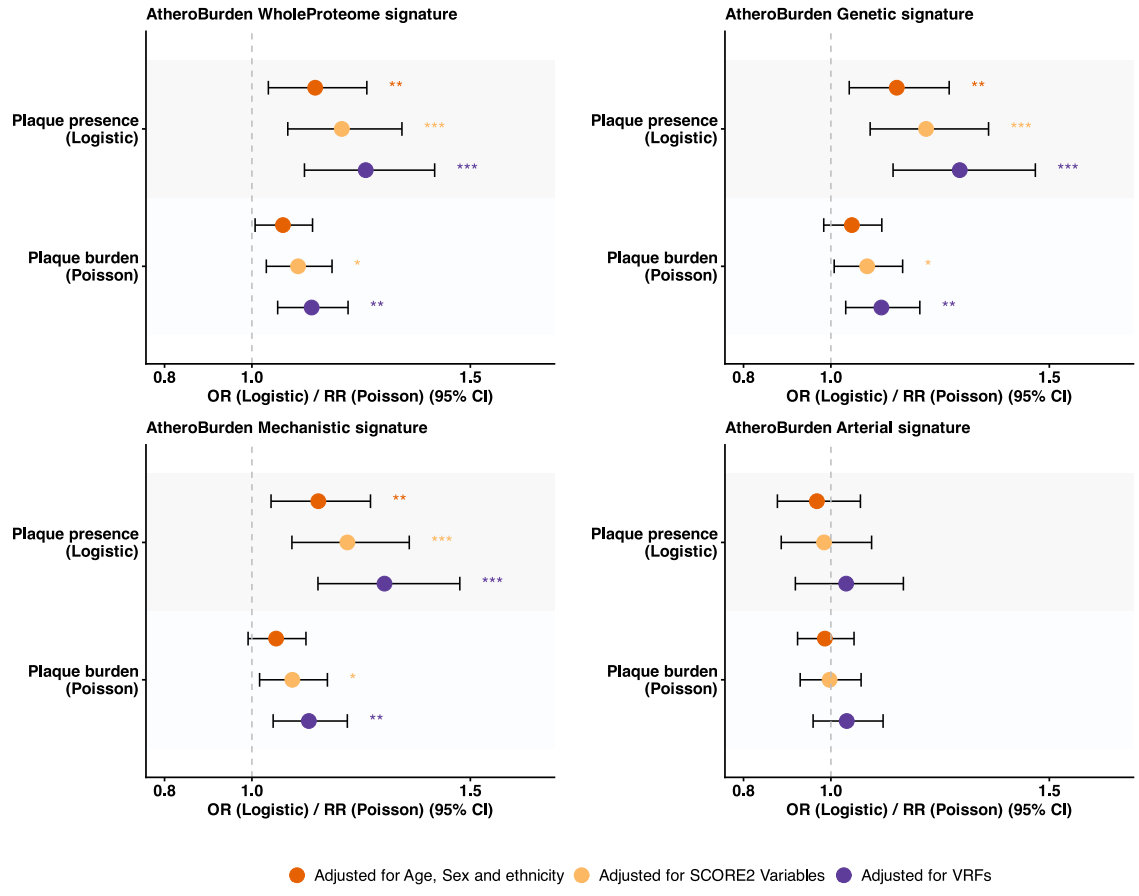

Figure 3 Associations between AtheroBurden signatures at baseline and ultrasound-defined carotid plaque presence and burden over follow-up in the UKB (n = 1712). Forest plots illustrating the associations between four AtheroBurden scores and two measures of atherosclerosis: plaque presence (assessed by logistic regression) and plaque burden (assessed by Poisson regression). Results are presented as ORs for plaque presence and RRs for plaque burden, each with corresponding 95% CIs. The grey dashed vertical line at 1.0 represents the null hypothesis of no association. Effect estimates are presented with 95% CIs under hierarchical adjustment models: demographic factors (age, sex, and ethnicity background), SCORE2 variables (age, sex, TC, HDL-C, SBP, and smoking status), and VRFs (age, sex, ethnicity background, SBP, BMI, smoking and drinking status, LDL-C, triglycerides, eGFR, glycated HbA1c, diabetes, hypertension, and family history of CVD). P-values were obtained from two-sided Wald tests for the AtheroBurden signatures in the corresponding logistic regression (plaque presence) and Poisson regression (plaque burden) models. Statistical significance after FDR adjustment is denoted by asterisks: *P < 0.05, **P < 0.01, ***P < 0.001. OR, odds ratio; RR, rate ratio.

was not performed at baseline in UKB (2006–10), we analysed 1712 participants who had baseline proteomic measurements and underwent carotid ultrasound at the first follow-up visit (starting in 2014). Using a deep learning model that we had previously developed,[59] we found 717 participants to have evidence of carotid atherosclerosis, of whom 222 participants had ≥2 plaques (see Supplementary material online, *Table S12*). In Logistic and Poisson regression models adjusted for age, sex, ethnic background, and VRFs, the WholeProteome, Genetic, and Mechanistic signatures were significantly associated with plaque presence and burden, whereas the Arterial signature showed no association (*Figure 3*; Supplementary material online, *Table S13*).

3.4 Prospective associations of AtheroBurden signatures with incident cardiovascular events

To examine the hypothesis that derived signatures capture presence and burden of atherosclerosis among individuals without a history of CVD, we next assessed prospective associations between the AtheroBurden signatures and incident MACE (composite of non-fatal AMI, non-fatal IS, and CV death) in an independent validation cohort of 41 200 participants (median age 58 years, 56% female, Supplementary material online, *Table S14*). During a median follow-up of 13.7 years, 2921 incident MACEs were documented. All four scores were consistently associated with incident MACE (*Figure 4A*) in Cox regression models adjusted for age, sex and ethnicity background, SCORE2 variables (age, sex, TC, HDL-C, SBP, and smoking status), as well as a more comprehensive list of demographic and VRFs (age, sex, ethnicity background, SBP, BMI, smoking and drinking status, LDL-C, triglycerides, eGFR, HbA1c, diabetes, hypertension, and family history of CVD). In the fully-adjusted models, the HR for MACE per standard deviation increase in the proteomic signatures ranged between 1.50 for the Arterial (95% CI: 1.44–1.57, $P = 5.9 \times 10^{-69}$) and Mechanistic signature (95% CI: 1.43–1.58, $P = 2.1 \times 10^{-57}$) and 1.58 for the Genetic (95% CI: 1.50–1.65, $P = 1.9 \times 10^{-80}$) and 1.57 for WholeProteome signature (95% CI: 1.50–1.65, $P = 3.9 \times 10^{-76}$, Supplementary material online, *Table S15*). All four signatures were significantly

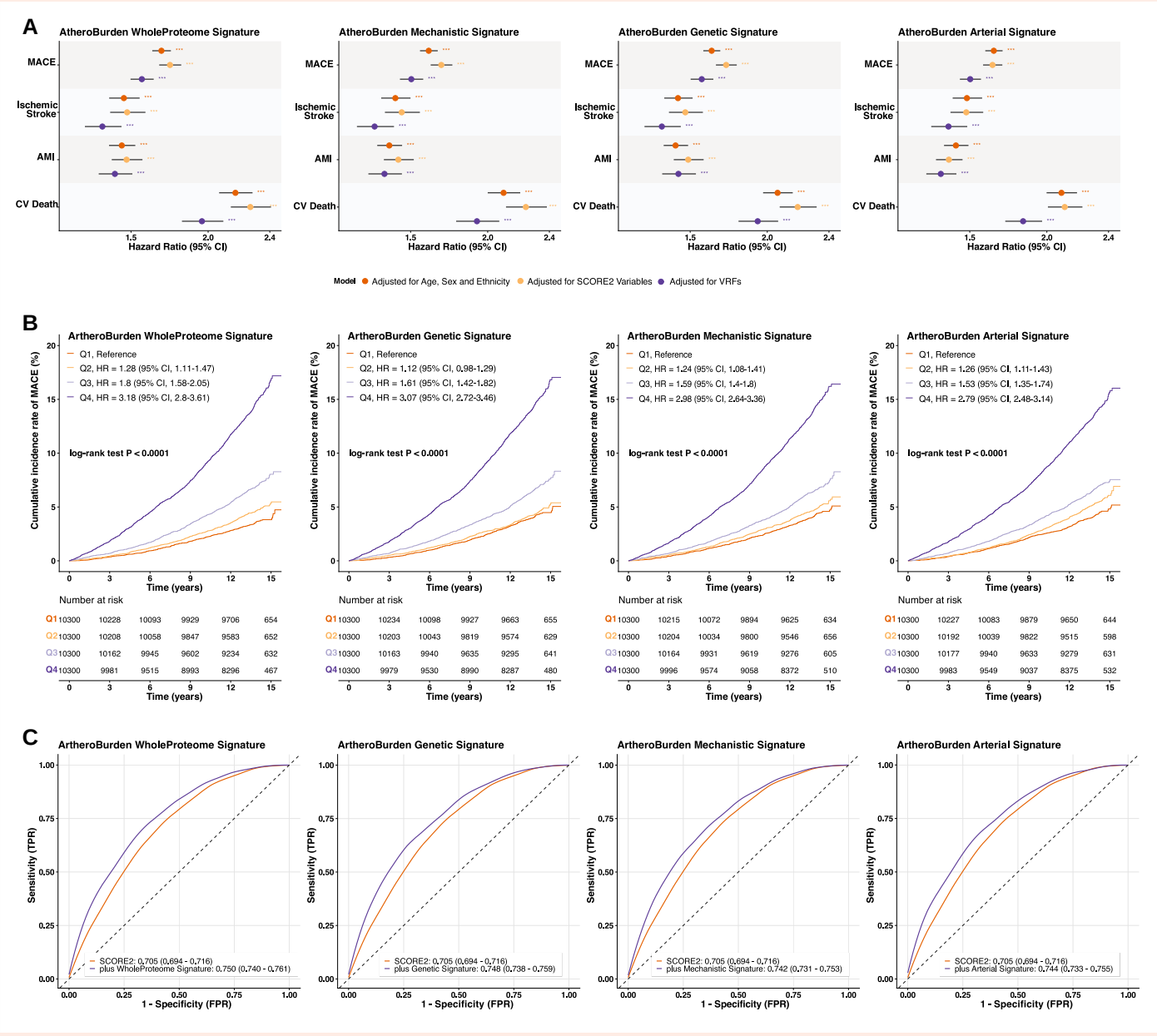

Figure 4 Associations of AtheroBurden scores with future cardiovascular risk in the UKB (n = 41 200). (*A*) Multivariable Cox regression analyses demonstrating associations between AtheroBurden scores and cardiovascular outcomes (MACE and its components—ischaemic stroke, AMI, and CV death). Effect estimates are presented with 95% CIs under hierarchical adjustment models: demographic factors (age, sex, and ethnicity background), SCORE2 variables (age, sex, TC, HDL-C, SBP, and smoking status), and VRFs (age, sex, ethnicity background, SBP, BMI, smoking and drinking status, LDL-C, triglycerides, eGFR, glycated HbA1c, diabetes, hypertension, and family history of CVD). *P*-values were obtained from two-sided Wald tests for the AtheroBurden signatures in Cox proportional hazards models. Statistical significance after FDR adjustment is denoted by asterisks: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. (*B*) Kaplan–Meier curves for the cumulative incidence of MACE stratified by quartiles of AtheroBurden signature. Population risk gradients are illustrated through colour-stratified quartiles (Q4: purple; Q1: orange), with HRs adjusted for SCORE2 variables. Risk tables quantify the at-risk population across follow-up intervals. Between-group differences were assessed using a two-sided log-rank test. (*C*) Time-dependent ROC evaluating discriminatory performance for predicting cardiovascular risk over a 10 year follow-up period. The orange curve represents the SCORE2 model alone, while the purple curve represents SCORE2 combined with AtheroBurden signatures. The enhancement in risk discrimination is quantified through comparative area under the curve metrics with corresponding 95% CIs. Q1/Q4, Quartile 1/Quartile 4.

associated with all three MACE components, but they showed consistently stronger associations with cardiovascular death than AMI and IS (*Figure 4*; Supplementary material online, *Table S15*). Stratifying the AtheroBurden scores by quartiles, we found strong dose–response relationships, with MACE risk with cumulative incidence of 16–17.2% in the highest (Q4) vs. 4.7–5.2% in the lowest quartiles (Q1) at the end of the 16 year follow-up (*Figure 4B*). The HR for Q4 vs. Q1 following adjustments for the full list of VRFs ranged from 2.79 (95% CI: 2.48–3.14) for the Arterial signature to 3.18 (95% CI: 2.8–3.61) for the WholeProteome signature.

Table 2 Incremental discrimination and reclassification improvement for predicting MACE with the addition of AtheroBurden signatures

Internal validation dataset (*N* = 41 200; 2921 incident MACE cases)				10 year follow-up (1770 incident MACE cases)	
Model	**C-index (95% CI)**	**ΔC-index (vs. SCORE2)**	***P*-value**	**cfNRI (95% CI)**	**IDI (95% CI)**
SCORE2	0.697 (0.686–0.708)	Ref	–	Ref	Ref
SCORE2 + AtheroBurden WholeProteome	0.741 (0.731–0.752)	0.044	5.49E−29	0.179 (0.155–0.200)	0.019 (0.015–0.024)
SCORE2 + AtheroBurden Genetic	0.739 (0.729–0.750)	0.042	2.35E−28	0.185 (0.159–0.208)	0.019 (0.015–0.023)
SCORE2 + AtheroBurden Mechanistic	0.733 (0.722–0.744)	0.036	2.46E−23	0.171 (0.148–0.194)	0.015 (0.011–0.019)
SCORE2 + AtheroBurden Arterial	0.735 (0.724–0.746)	0.038	6.86E−23	0.170 (0.144–0.196)	0.020 (0.015–0.025)
Analysis in males (*N* = 18 057; 1813 incident MACE cases)				10 year follow-up (1142 incident MACE cases)	
SCORE2	0.675 (0.661–0.690)	Ref	–	Ref	Ref
SCORE2 + AtheroBurden WholeProteome	0.734 (0.720–0.747)	0.058	9.35E−23	0.229 (0.198–0.257)	0.031 (0.025–0.040)
SCORE2 + AtheroBurden Genetic	0.734 (0.720–0.748)	0.059	1.27E−24	0.231 (0.202–0.266)	0.033 (0.026–0.042)
SCORE2 + AtheroBurden Mechanistic	0.728 (0.714–0.742)	0.053	4.42E−21	0.229 (0.196–0.259)	0.028 (0.021–0.035)
SCORE2 + AtheroBurden Arterial	0.729 (0.714–0.743)	0.053	8.52E−20	0.208 (0.177–0.236)	0.031 (0.025–0.039)
Analysis in females (*N* = 23 143; 1108 incident MACE cases)				10 year follow-up (628 incident MACE cases)	
SCORE2	0.722 (0.703–0.740)	Ref	–	Ref	Ref
SCORE2 + AtheroBurden WholeProteome	0.749 (0.732–0.767)	0.027	1.06E−07	0.144 (0.102–0.182)	0.009 (0.006–0.014)
SCORE2 + AtheroBurden Genetic	0.749 (0.732–0.767)	0.027	1.95E−07	0.165 (0.128–0.204)	0.009 (0.006–0.015)
SCORE2 + AtheroBurden Mechanistic	0.745 (0.727–0.763)	0.023	2.28E−06	0.153 (0.109–0.188)	0.007 (0.004–0.012)
SCORE2 + AtheroBurden Arterial	0.745 (0.727–0.763)	0.023	1.00E−05	0.147 (0.104–0.189)	0.011 (0.007–0.018)

Comparison of model discrimination (C-index and ΔC-index) and reclassification (cfNRI and IDI) when adding various AtheroBurden signatures to SCORE2 for predicting MACEs. Results include internal validation and subgroup analyses by sex. All comparisons use SCORE2 as Ref.
Ref, reference.

Adding the AtheroBurden signatures to baseline SCORE2 led to significantly improved discrimination for future MACE risk, as indicated by increases in the C-indices (*Table 2*). The WholeProteome signature exhibited the largest improvement in discrimination, increasing the C-index by 0.04 (from 0.70 to 0.74; $P = 5.49 \times 10^{-29}$). These improvements remained robust in sex-stratified analyses, yielding an increase of up to 0.06 in the C-index among males. Testing discrimination changes in 10 year risk, against which SCORE2 is validated, further supported significant improvements (timeROC for SCORE2 0.71 vs. 0.75 when adding the AtheroBurden WholeProteome signature, $P = 1.76 \times 10^{-32}$, *Figure 4C*). Incorporating AtheroBurden signatures also led to improvements in calibration, as indicated by improved alignment between predicted and observed risks (see Supplementary material online, *Extended Figure S7*), as well as in NRI metrics (cfNRI and IDI) for both the 10 year and total follow-up periods ($P < 0.001$ for all comparisons; *Table 2*). At established clinical decision risk thresholds (7.5 and 10%), addition of AtheroBurden signatures to SCORE2 led to improved reclassification of study participants to the right risk category. For example, the WholeProteome signature improved net reclassification of 11.8% of study participants (95% CI: 9.4–14.6%) at the 10% risk threshold, while the Genetic signature yielded a 10% improvement (95% CI: 7.5–12.9%) at the 7.5% threshold (see Supplementary material online, *Extended Figure S8*).

3.5 Temporal trajectories of AtheroBurden signatures and their clinical correlates

To examine whether AtheroBurden signatures change over time and track with disease progression, we analysed 1210 UKB participants with serial proteomic assessments at follow-up visits (Instance 2 starting in 2014 or Instance 3 starting in 2019). We first assessed associations of baseline VRFs with annual changes in signature scores and then examined whether individuals who subsequently experienced MACE showed different progression patterns compared with event-free participants.

Compared with participants with proteomics profiling at a single timepoint, participants with serial assessments were significantly younger (median age at recruitment 49 vs. 58 years) and had a substantially lower burden of VRFs (see Supplementary material online, *Table S16*). Because follow-up proteomic assessments at these time points were limited to an earlier version of Olink Explore—covering ~50% of the proteins in the newer version—we generated restricted signatures using the overlapping subset of 1459 proteins (see Supplementary material online, *Table S17*). These restricted signatures correlated highly with the full signatures derived at baseline (Instance 0), demonstrating robust signal preservation ($R = 0.91$–0.94, all $P < 2.2 \times 10^{-19}$; *Figure 5A*).

When stratified by baseline cardiovascular risk categories, we found individuals in higher baseline cardiovascular risk strata (10 year SCORE2 risk: <2.5, 2.5–5, 5–7.5, and >7.5%) to exhibit steeper annual increases in all four AtheroBurden signatures (*Figure 5B*). For example, changes in the Genetic AtheroBurden signature ranged from a decrease of 0.020 SD per year (95% CI: −0.026 to −0.015, $P = 3.61 \times 10^{-13}$) in the lowest baseline risk category (<2.5% 10 year risk) to an increase of 0.085 SD per year (95% CI: 0.076–0.094; $P = 1.15 \times 10^{-13}$) in the highest risk category (>7.5%). Furthermore, to determine whether AtheroBurden signature progression was specifically associated with clinical outcomes, we tested the progression patterns of individuals who went on to experience incident MACE during follow-up using linear mixed-effects models, which revealed significantly different

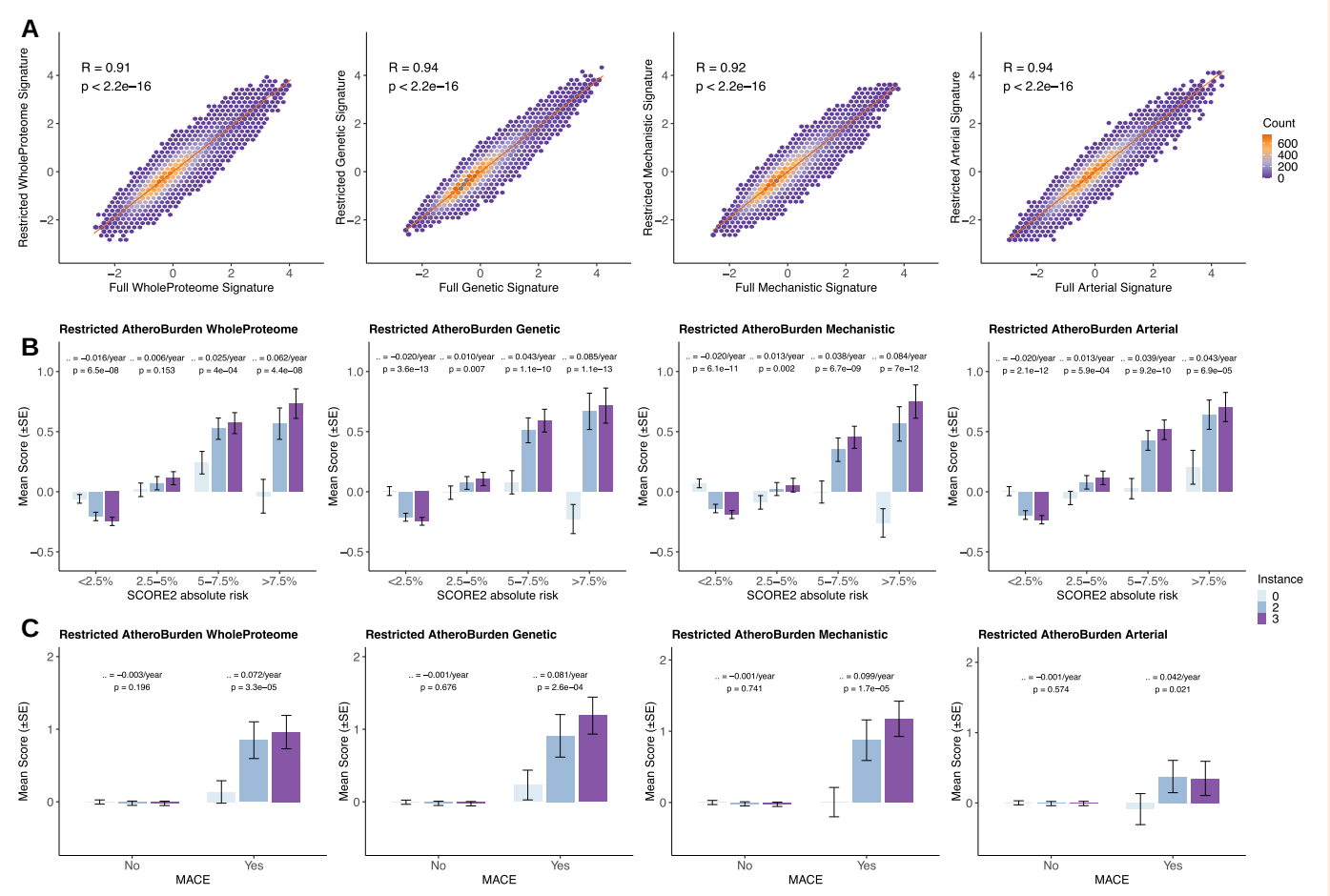

Figure 5 Serial changes of AtheroBurden Scores by baseline cardiovascular risk and incident cardiovascular events in the UKB (n = 1210). (*A*) Baseline correlation of restricted and full proteomic signatures. Restricted proteomic signatures were derived at baseline (Instance 0) using 1459 proteins common across both available measurement platforms (Olink Explore 1536 and Explore 3072). Scatter plots with hexagonal binning illustrate correlations between restricted and corresponding full AtheroBurden signatures (WholeProteome, Genetic, Mechanistic, and Arterial). Pearson correlation coefficients (*R*) and associated *P*-values are displayed for each signature. *P*-values were obtained using a two-sided Pearson correlation test. (*B*) Longitudinal trajectories stratified by baseline SCORE2 risk categories. Temporal evolution of AtheroBurden scores stratified by baseline SCORE2 risk categories (<2.5, 2.5–5, 5–7.5, and >7.5%). Bars represent mean values at three time points (Instances 0, 2, and 3), with error bars indicating SE. Annual progression rates (β) and corresponding *P*-values were derived from linear mixed-effects models. *P*-values correspond to two-sided *t*-tests. (*C*) Temporal evolution of AtheroBurden scores stratified by incident MACEs. Mean AtheroBurden scores at three time points (Instances 0, 2, and 3) are presented separately for participants without and with subsequent MACE events (denoted as 'No' and 'Yes', respectively). Error bars represent SE. Annual progression rates (β) and *P*-values were derived from linear mixed-effects models. *P*-values correspond to two-sided *t*-tests. SE, standard error.

trajectories (*Figure 5C*). Specifically, we found progression of AtheroBurden signatures to be restricted to individuals who experienced MACE during follow-up. Annual progression coefficients in MACE-positive participants ranged from $\beta = 0.042$ (95% CI: 0.007–0.076; $P = 0.021$) for the Arterial signature to $\beta = 0.099$ (95% CI: 0.058–0.139; $P = 1.73 \times 10^{-5}$) for the Genetic signature, while event-free participants exhibited no significant progression.

3.6 External validation of AtheroBurden signatures in KORA cohorts

Finally, to externally validate our findings, restricted versions were applied to the population-based KORA S4 (n = 1361) and KORA-Age1 (n = 796) prospective cohort studies (see Supplementary material online, *Table S18*). Because protein quantification in KORA cohorts was limited to cardiovascular and inflammation panels, restricted signatures were constructed using available overlapping proteins: 232 for WholeProteome, 44 for Genetic, 146 for Mechanistic, and 30 for Arterial signatures in KORA S4 and 242 for WholeProteome, 45 for Genetic, 147 for Mechanistic, and 30 for Arterial signatures in KORA-Age1. There were moderate to strong correlations between the KORA-adapted signatures and the full AtheroBurden signatures in UKB (R = 0.56–0.75, all $P < 2.2 \times 10^{-16}$) in both cohorts (*Figure 6A* for S4; Supplementary material online, *Extended Figure S9A* for Age1). Compared with UKB, KORA participants were older (median age: S4, 63 years; Age1, 76 years) with a higher prevalence of hypertension (S4: 54%; Age1: 75% vs. UKB: 25%) and more balanced sex distribution (S4: 50% female; Age1: 53% female, *Table 1*). Over a median follow-up of 15.1 years in S4 and 6.8 years

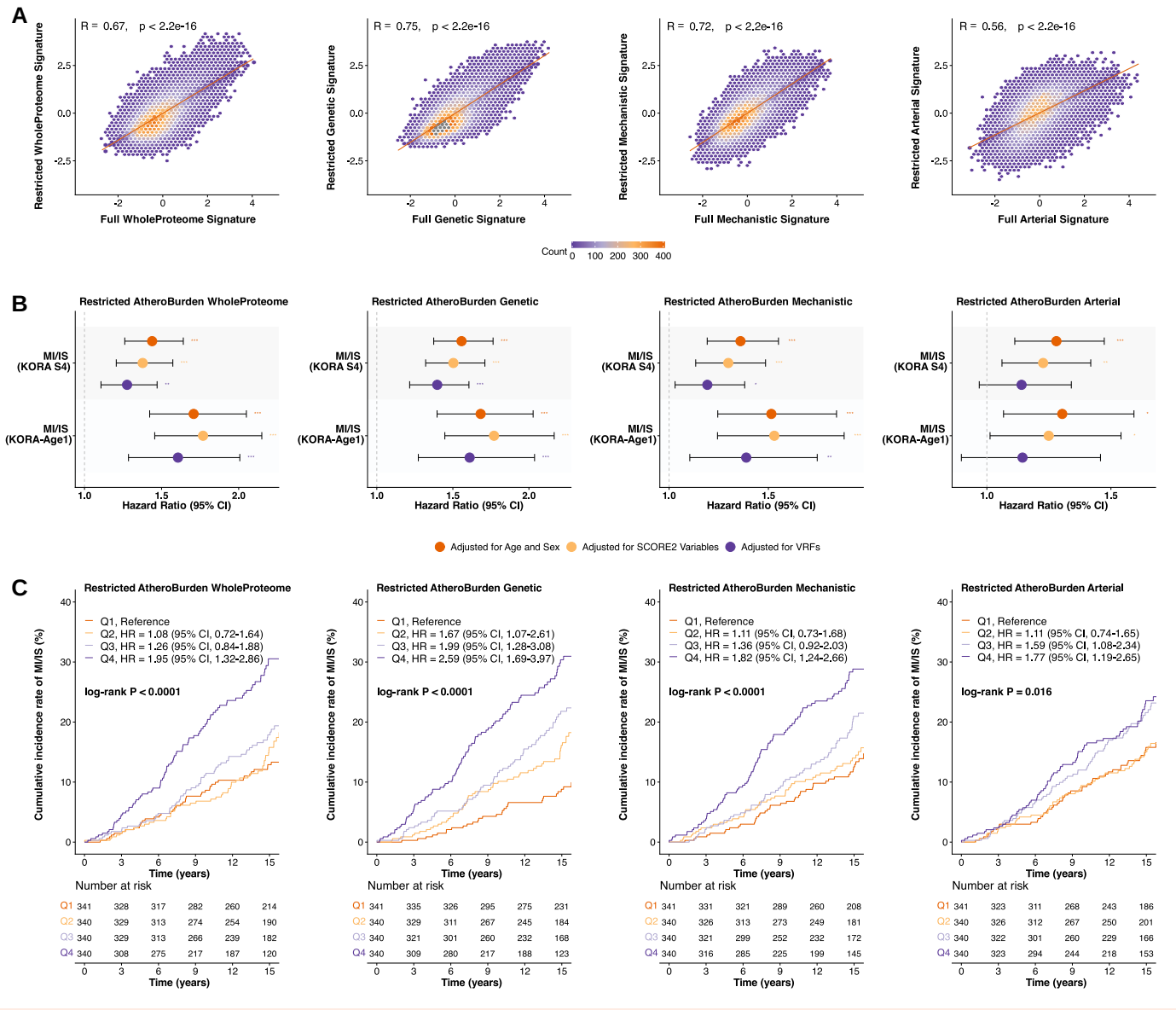

Figure 6 Associations of AtheroBurden signatures with future cardiovascular risk in the KORA S4 ($n = 1361$) and KORA-Age1 ($n = 796$) cohorts. (*A*) Correlation between restricted (KORA S4) and full proteomic signatures in the UKB baseline cohort. Restricted signatures (KORA S4) were derived using only proteins quantifiable across all measurement platforms. Hexagonal binning scatter plots demonstrate correlations between restricted and corresponding full signatures with Pearson correlation coefficients (*R*) and associated two-sided *P*-values. (*B*) Forest plots present HRs and 95% CIs for restricted AtheroBurden scores across the KORA S4 and Age1 cohorts. HRs are shown for three adjustment models: demographic factors (age and sex), SCORE2 variables (TC, HDL-C, SBP, and smoking status), and VRF model The VRF model included age, sex, SBP, BMI, smoking and drinking status, LDL-C, triglycerides, eGFR, glycated HbA1c, diabetes, and hypertension status in both cohorts. Additionally, in KORA S4, parental history of CVD and country of birth (born in Germany) were included. The grey dashed line indicates an HR of 1.0 (no association). *P*-values were obtained from two-sided Wald tests for the AtheroBurden signatures in Cox proportional hazards models. Asterisks denote FDR-adjusted *P*-values within each cohort: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. (*C*) Kaplan–Meier curves showing cumulative incidence rates of MI/stroke stratified by quartiles of four restricted AtheroBurden signatures in KORA S4. The HRs displayed are adjusted for SCORE2 variables. Risk tables are provided below each plot, and log-rank test *P*-values are displayed for group comparisons.

in Age1, 233 and 108 participants were diagnosed with an MI or IS, respectively.

In age- and sex-adjusted models, we found all four AtheroBurden signatures to be associated with the risk of MI or IS in both cohorts (*Figure 6B*; Supplementary material online, *Table S19*), with the Mechanistic and WholeProteome signatures demonstrating significant associations after adjustment for the full set of VRFs. Kaplan–Meier analyses demonstrated that participants in the highest signature quartile (Q4) exhibited significantly elevated cumulative incidence of cardiovascular events compared with lower quartiles in KORA S4 (log-rank $P < 0.05$ for all signatures, *Figure 6C*). After adjusting for SCORE2 variables, HRs for Q4 relative to Q1 ranged from 1.77 (95% CI: 1.19–2.65) for the Arterial signature to 2.59 (95% CI: 1.69–3.97) for the Genetic signature. Similar patterns

were observed in KORA-Age1 (see Supplementary material online, *Extended Figure S9B*). Sensitivity analyses excluding overlapping participants in S4 and Age1 showed similar results, though with limited statistical power to draw definitive conclusions (see Supplementary material online, *Extended Figure S9C* and *Table S20*). Similar to the UKB, adding the AtheroBurden signatures to the baseline SCORE2 model improved discrimination of MI or IS in both KORA S4 (ΔC-index range: 0.010–0.034) and KORA-Age1 (ΔC-index range: 0.020–0.050; Supplementary material online, *Table S21*).

4. Discussion

In this study, by leveraging large-scale population-based data, we constructed four plasma proteomic signatures that (i) discriminated between presence and absence of clinically diagnosed atherosclerotic disease, (ii) showed a dose–response relationship with the number of vascular beds affected by atherosclerosis, (iii) correlated with imaging-defined carotid plaque burden, (iv) strongly predicted future risk of cardiovascular events in disease-free individuals, and (v) longitudinally changed according to baseline cardiovascular risk and future MACE occurrence. Our findings demonstrate the utility of ML-derived signatures of plasma proteomics for assessing atherosclerosis burden and estimating cardiovascular risk in disease-free individuals.

Our findings extended prior plasma proteomics work in two complementary directions. On the one hand, several large-scale proteomic risk scores have been optimized primarily for incident cardiovascular event prediction, generally yielding modest but statistically significant improvements beyond established clinical risk models (ΔC-index ~0.01–0.05; NRI ~4–14%).[35,36,38,41] On the other hand, proteomics has also been linked to subclinical atherosclerosis phenotypes, including carotid plaque burden, carotid intima-media thickness, and coronary artery calcium, although most studies have focused on single vascular beds or targeted protein panels, with incremental performance varying across settings.[42–48] Our framework bridges these approaches by explicitly targeting systemic atherosclerotic burden across multiple vascular beds and linking this burden to subsequent vascular events through four complementary signatures, with external validation and robustness analyses under constrained protein coverage. In this context, our data provide convergent evidence supporting the potential utility of the AtheroBurden signatures as circulating biomarkers of atherosclerosis burden. First, all four proteomic signatures demonstrated strong discriminative performance in identifying individuals with a history of atherosclerotic disease and correlated with disease burden, as reflected by the number of affected vascular beds. Second, we found the signatures to be associated with plaque presence and count in carotid ultrasound —an imaging-based surrogate of subclinical atherosclerosis—further reinforcing their relevance to underlying disease biology. Third, among asymptomatic individuals without evidence of atherosclerotic disease, higher values of all four signatures were associated with substantially increased risks of adverse cardiovascular events in UKB and KORA. Individuals who went on to develop cardiovascular events are expected to have a higher burden of atherosclerosis at baseline. Persistence of the associations after adjustment for traditional vascular risk factors suggests that the proteomic signatures provide biological information beyond standard risk metrics. Moreover, sustained discrimination after excluding established biomarkers such as natriuretic-peptide and apolipoprotein-related analytes indicates that predictive information is distributed across multiple proteins rather than dominated by a single established marker. Notably, the observed ΔC-index improvement over SCORE2 (+0.044) approximates the pooled gain of adding coronary artery calcium in a meta-analysis (+0.036; 95% CI 0.020–0.052),[76] supporting a scalable, repeatable, radiation-free blood-based readout of systemic atherosclerotic burden that aligns with imaging-defined plaque burden. Fourth, longitudinal data across three serial time points over a median follow-up of 12.5 years revealed that AtheroBurden scores track with disease progression. Steeper annual increases were observed among individuals with greater baseline vascular risk and among those who subsequently experienced major cardiovascular events, consistent with the trajectory of atherogenesis. Collectively, these proteomic signatures could act as dynamic, non-invasive biomarkers of atherosclerotic burden. Importantly, the AtheroBurden framework is designed to be interpreted as an integrated composite score rather than through individual protein fluctuations. The disease-relevant signal was preserved even with substantially reduced protein coverage: the biologically constrained Genetic and Mechanistic signatures performed comparably with the WholeProteome signature. Robustness was further supported under real-world technical constraints with lower protein coverage in longitudinal UKB analyses and through external validation in KORA cohorts. Together, these findings indicate that the predictive information captured by the AtheroBurden framework does not depend on exhaustive protein coverage. Nonetheless, prospective validation in independent cohorts with integrated vascular imaging and proteomic profiling will be essential to confirm their utility as biomarkers of subclinical disease and progression.

We developed four distinct proteomic signatures, each comprising proteins with varying relevance to atherosclerosis. Beyond the Arterial signature, the Mechanistic, Genetic, and WholeProteome signatures demonstrated comparable performance in detecting atherosclerotic disease and predicting future MACE events. Our approach of not relying solely on the whole-proteome panel aimed to reduce the influence of proteins whose circulating levels may reflect secondary effects of tissue ischaemia rather than atherosclerosis progression. Investigating the top-ranked proteins in each panel provides insights into the distinct biological signals captured by our signatures. In the artery-enriched panel, highly-ranked proteins were specific to cardiovascular tissues, such as neurotrophic receptor tyrosine kinase 3, implicated in cardiac remodelling,[77] leptin, involved in energy homeostasis[78] and linked to subclinical atherosclerosis,[79] and angiopoietin-2 (ANGPT2), which has shown prognostic relevance in peripheral artery disease[80] and intracranial stenotic lesions.[81] Additional high-ranking proteins are linked to extracellular matrix remodelling, growth factor signalling, and inflammation included integrin alpha-11 (ITGA11), insulin-like growth factor-binding protein 3 (IGFBP3), transforming growth factor-beta-induced protein (TGFBI), and lipopolysaccharide-bingding protein (LBP). These proteins have established roles in cardiovascular pathology: ITGA11 in CAD susceptibility and cardiac fibroblast differentiation,[82,83] IGFBP3 in atherosclerotic plaque stability modulation,[84] and TGFBI and LBP in macrophage-mediated inflammatory responses.[85–87] Consistent with these functional annotations, leptin tracked adiposity and LBP tracked systemic inflammation in UKB (see Supplementary material online, *Extended Figure S6* and *Table S11*), supporting that the artery-enriched panel captures vascular/tissue-relevant and inflammatory biology. In contrast, the Mechanistic, Genetic, and WholeProteome panels predominantly prioritized systemic cardiovascular markers, including NT-proBNP and REN, as well as cholesterol-associated proteins like PCSK9 and APOC1, potentially reflecting end-organ dysfunction and high cardiovascular risk. In line with this interpretation, NT-proBNP covaried with age and kidney function, while lipid-associated proteins covaried with clinical lipid measures (see Supplementary material online, *Extended Figure S6* and *Table S11*), providing additional context for the biological signals captured by these signatures. Notably, lipoprotein-associated phospholipase A2 (Lp-PLA2; PLA2G7) was also prioritized among the top-ranked proteins across our

signatures. Given that Lp-PLA2 circulates predominantly on LDL-C/ApoB-containing lipoproteins, PLA2G7 may partly reflect the atherogenic lipoprotein axis. This interpretation is particularly relevant in the UKB Olink context, where the ApoB analyte shows poor concordance with clinical ApoB (Spearman $\rho = 0.08$) and no association with 10 year incident major cardiovascular events.[88]

If validated, plasma proteomic signatures of atherosclerosis could have two key translational applications. First, they could enable monitoring of atherosclerosis progression and cardiovascular risk in the context of primary prevention. Circulating biomarkers are scalable in primary care settings and could complement imaging by identifying individuals with a higher underlying atherosclerotic burden, prioritizing them for confirmatory imaging, tracking cardiovascular risk over time, and monitoring responses to preventive interventions. While current proteomics assays are costly, the development of targeted protein panels that capture most of the variance in the full signatures could reduce costs and promote clinical implementation. Second, proteomic signatures may have utility in drug development, both as for patient stratification tools and as surrogate endpoints of efficacy in trials of atheroprotective treatments. Our preliminary analysis in a small subset with serial measurements suggests that these signatures may reflect atherosclerosis progression over time, but whether they respond to treatment effects remains uncertain. In *post-hoc* analyses of two Phase 3 trials testing the glucagon-like peptide-1 receptor agonist (GLP-1RA) semaglutide, randomization to treatment vs. placebo led to significant reductions in proteomic signatures associated with MACE risk.[89] In the absence of scalable non-imaging-based endpoints for atherosclerosis, proteomic signatures may offer a promising surrogate endpoint for early-phase trials. Incorporating proteomic profiles into Phase 3 cardiovascular outcomes trials could enable evaluation of whether such signatures correlated with treatment effects on risk reduction at an individual level.

Our study has several limitations. First, the use of clinical diagnoses as proxies for atherosclerotic disease may have biased our signatures towards higher disease burden. As clinical diagnoses typically reflect stenotic atherosclerotic lesions, it remains unknown whether our signatures also capture very early atherosclerotic changes. Although supplementary analyses incorporating carotid plaque phenotyping were reassuring, the ultrasound assessments in UKB were only performed 8 years after the initial proteomic profiling. Future studies that integrate comprehensive vascular imaging with contemporary proteomic profiling could enable more precise phenotyping for model development. Second, external replication in KORA cohorts was constrained by reduced proteomic coverage (276 vs. 2920 proteins) and differences in cohort characteristics—KORA participants were older with a higher prevalence of hypertension, potentially enriching for individuals with greater atherosclerotic burden. To preserve the independence of external validation, we did not refit/retrain the models in KORA; instead, we applied UKB-developed, panel-adapted implementations based on overlapping proteins. Despite these differences, restricted signatures based on overlapping proteins showed moderate to strong correlations with the full signatures ($R = 0.56–0.75$) and remained significantly associated with future MI and stroke events, supporting the robustness and generalizability of the multi-protein strategy even when some canonical biomarkers are missing. While the KORA S4 cohort provided sufficient power for robust Cox regression analyses, replication in the smaller and older KORA-Age1 cohort was limited by reduced sample size and event count. Third, repeated proteomic measurements were available only in a small, relatively healthy subset of 1210 UKB participants enrolled in the COVID-19 repeat-imaging study, with only 10 of 25 MACE occurring after Instance 2, limiting statistical power for event-stratified trajectory inference. After addressing proteomic coverage differences between time points (1463 proteins in Instances 2/3 vs. 2923 in Instance 0) using restricted signatures ($R = 0.91–0.94$), we observed divergent trajectories between participants with and without incident MACE, as well as across SCORE2 risk categories. However, given reported associations between COVID-19 and cardiovascular risk,[90,91] these trajectory changes may partly reflect COVID-related inflammatory responses rather than atherosclerosis progression alone. Accordingly, these findings require cautious interpretation and replication in independent cohorts with serial proteomic measurements. Fourth, the predominantly European ancestry of the study populations may limit generalizability to other ethnic groups. Future studies in more diverse cohorts are needed to assess the transferability and robustness of the identified signatures across ancestries. Fifth, the selection of proteins for the artery-enriched signature was based on gene expression profiles from the GTEx database,[69] which includes bulk tissue samples from coronary, aortic, and tibial arteries of donors aged 20–71 years who died due to any cause. As these tissues were not selected specifically for atherosclerosis involvement, the resulting protein panel may lack specificity for proteins derived from atherosclerotic plaques. This limitation could partly explain the relatively modest performance of the artery-enriched signature in predicting future MACE and carotid plaque presence compared with the other protein scores. Future studies leveraging proteomic or transcriptomic data directly from human atherosclerotic plaque tissue may yield more atherosclerosis-specific signatures. Sixth, while NPPB, NT-proBNP, and REN have published orthogonal validation showing good concordance of Olink-based measurements with conventional assays,[92–94] most other key proteins in our signatures lack such validation. Future work is needed to confirm absolute quantification and clinical transferability with reference methods.

In conclusion, plasma proteomic signatures can capture atherosclerosis burden, improving cardiovascular risk prediction in asymptomatic individuals. If replicated in cohorts bridging extensive vascular phenotyping with proteomic profiling, our results suggest that the circulating proteome could serve as an accessible readout of systemic atherosclerotic burden that complements imaging-based plaque burden assessment. This approach could enable broader implementation of screening and prevention strategies for CVD.

Supplementary material

Supplementary material is available at *Cardiovascular Research* online.

Conflict of interest: M.K.G. reports consulting fees from Tourmaline Bio, Inc., Pheiron GmbH, Dexcel Pharma Technologies Ltd., Novartis AG and Gerson Lehrman Group, Inc. V.L.M. owns stock or stock options in Abbott Labs, Abbvie, Amgen, Baxter International, Boston Scientific, Cardinal Health, Eli Lilly, GE Healthcare, Ionetix, Johnson & Johnson, Merck, Organon, Pfizer, Viatris, and Zimvie. V.L.M. has received research grants and consulting fees from Siemens Medical Imaging and consulting fees from INVIA Medical Imaging Solutions, Jubilant Draximage, and Credence Management Solutions. The other authors have nothing to disclose.

Funding

This work was funded by the Fritz-Thyssen Foundation (grant ref. 10.22.2.024MN), the German Research Foundation (DFG; Emmy Noether grant GZ: GE 3461/2-1, ID 512461526; Excellence Strategy grant within the framework of the Munich Cluster for Systems Neurology EXC 2145 SyNergy, ID 390857198; and the Collaborative Research Center 1744 [ID 548585053]), the Gemeinnützige Hertie-Stiftung (Hertie Network of Excellence in Clinical Neuroscience, ID P1230035). L.Z. would like to thank the

China Scholarship Council for the financial support (File No. 202306170051). S.N.G is supported by U.S. Departement of Veterans Affairs (VA MERIT 1I01CX002560), Taubman Medical Research Institute (Wolfe Scholarship), and NIH/NHLBI R01HL150392. M.D. has received funding from the Vascular Dementia Research Foundation. Z.Y. is supported by the National Human Genome Research Institute (R00HG012956). V.L.M. receives research support from the National Heart, Lung, and Blood Institute (R01HL136685), the National Institute on Aging (R01AG059729), and the National Institute of Diabetes and Digestive and Kidney Diseases (U01DK123013) and is also supported by the Melvyn Rubenfire Professorship in Preventive Cardiology. The KORA study was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Data collection in the KORA study is done in cooperation with the University Hospital of Augsburg. OLINK proteomics measurements in KORA S4 were partly supported by the Helmholtz Institute for Metabolic, Obesity and Vascular Research—Project Initiative 2018 (HI-MAG).

Data availability

The coronary artery disease GWAS summary statistics are available through the GWAS Catalog (https://www.ebi.ac.uk/gwas/) (accession no. GCST90132315) and CARDIoGRAMplusC4D (https://www.cardiogramplusc4d.org/data-downloads/). The proteomic GWAS summary data from the UK Biobank Pharma Proteomics Project (UKB-PPP) can be accessed publicly via Synapse (https://www.synapse.org/Synapse:syn51365301). To protect patient confidentiality and ensure compliance with consent agreements, individual-level data and proteomic profiles from UK Biobank and KORA are available under controlled access. UK Biobank data can be accessed by approved researchers through the UK Biobank data access framework (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access), with the full dataset available on the Research Analysis Platform (https://www.ukbiobank.ac.uk/enable-your-research/research-analysis-platform). KORA datasets are available on reasonable request through a project agreement from KORA (https://helmholtz-muenchen.managed-otrs.com/external/). Requests should be sent to kora.passt@helmholtz-munich.de and are subject to approval by the KORA board. The codes used for this study are available on GitHub (https://github.com/DeepVasc-Lab/AtheroBurden_proteomics_signatures_ukb).

Ethical approval

Ethical approval for UK Biobank data use was obtained from the North West-Haydock Research Ethics Committee (REC reference: 16/NW/0274). This research was conducted under approved application numbers 151281. Access to individual-level data from KORA was granted under application number 20240709720 00087. KORA was approved by the Ethics Committee of the Bavarian Chamber of Physicians, Munich (KORA S4: EC No. 99186) and the Bavarian Medical Association (KORA-Age: EC No. 08094), with all participants providing written informed consent. The GWAS datasets used in this study are publicly available and do not require additional ethical approval.

References

1. Mensah GA, Fuster V, Murray CJL, Roth GA; Global Burden of Cardiovascular Diseases and Risks Collaborators. Global burden of cardiovascular diseases and risks, 1990-2022. *J Am Coll Cardiol* 2023;**82**:2350–2473.
2. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, Barengo NC, Beaton AZ, Benjamin EJ, Benziger CP, Bonny A, Brauer M, Brodmann M, Cahill TJ, Carapetis J, Catapano AL, Chugh SS, Cooper LT, Coresh J, Criqui M, DeCleene N, Eagle KA, Emmons-Bell S, Feigin VL, Fernández-Solà J, Fowkes G, Gakidou E, Grundy SM, He FJ, Howard G, Hu F, Inker L, Karthikeyan G, Kassebaum N, Koroshetz W, Lavie C, Lloyd-Jones D, Lu HS, Mirijello A, Temesgen AM, Mokdad A, Moran AE, Muntner P, Narula J, Neal B, Ntsekhe M, Moraes de Oliveira G, Otto C, Owolabi M, Pratt M, Rajagopalan S, Reitsma M, Ribeiro ALP, Rigotti N, Rodgers A, Sable C, Shakil S, Sliwa-Hahnle K, Stark B, Sundström J, Timpel P, Tleyjeh IM, Valgimigli M, Vos T, Whelton PK, Yacoub M, Zuhlke L, Murray C, Fuster V. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. *J Am Coll Cardiol* 2020;**76**:2982–3021.
3. Zaman S, Wasfy JH, Kapil V, Ziaeian B, Parsonage WA, Sriswasdi S, Chico TJA, Capodanno D, Colleran R, Sutton NR, Song L, Karam N, Sofat R, Fraccaro C, Chamié D, Alasnag M, Warisawa T, Gonzalo N, Jomaa W, Mehta SR, Cook EES, Sundström J, Nicholls SJ, Shaw LJ, Patel MR, Al-Lamee RK. The Lancet Commission on rethinking coronary artery disease: moving from ischaemia to atheroma. *Lancet* 2025;**405**:1264–1312.
4. Hansson GK, Hermansson A. The immune system in atherosclerosis. *Nat Immunol* 2011;**12**:204–212.
5. Libby P. Inflammation in atherosclerosis. *Nature* 2002;**420**:868–874.
6. Global Cardiovascular Risk Consortium. Global effect of cardiovascular risk factors on lifetime estimates. *N Engl J Med* 2025;**393**:125–138.
7. Goh RSJ, Chong B, Jayabaskaran J, Jauhari SM, Chan SP, Kueh MTW, Shankar K, Li H, Chin YH, Kong G, Anand VV, Chan KA, Sukmawati I, Toh SA, Muthiah M, Wang J-W, Tse G, Mehta A, Fong A, Baskaran L, Zhong L, Yap J, Yeo KK, Hausenloy DJ, Tan JWC, Chao T-F, Li Y-H, Lim SL, Chan KH, Loh PH, Chai P, Yeo TC, Low AF, Lee CH, Foo R, Tan HC, Yip J, Rao S, Honda S, Yasuda S, Kajiya T, Goto S, Yan BP, Zhou X, Figtree GA, Mamas MA, Kim Y, Jeong Y-H, Kim MH, Park D-W, Park S-J, Richards AM, Chan MY, Lip GYH, Chew NWS. The burden of cardiovascular disease in Asia from 2025 to 2050: a forecast analysis for East Asia, South Asia, South-East Asia, Central Asia, and high-income Asia Pacific regions. *Lancet Reg Health West Pac* 2024;**49**:101138.
8. Joynt Maddox KE, Elkind MSV, Aparicio HJ, Commodore-Mensah Y, de Ferranti SD, Dowd WN, Hernandez AF, Khavjou O, Michos ED, Palaniappan L, Penko J, Poudel R, Roger VL, Kazi DS. Forecasting the burden of cardiovascular disease and stroke in the United States through 2050-prevalence of risk factors and disease: a presidential advisory from the American Heart Association. *Circulation* 2024;**150**:e65–e88.
9. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB Sr, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, Robinson JG, Schwartz JS, Shero ST, Smith SC, Sorlie P, Stone NJ, Wilson PWF. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2014;**63**:2935–2959.
10. D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;**117**:743–753.
11. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;**357**:j2099.
12. Daghem M, Bing R, Fayad ZA, Dweck MR. Noninvasive imaging to assess atherosclerotic plaque composition and disease activity: coronary and carotid applications. *JACC Cardiovasc Imaging* 2020;**13**:1055–1068.
13. Wust RCI, Calcagno C, Daal MRR, Nederveen AJ, Coolen BF, Strijkers GJ. Emerging magnetic resonance imaging techniques for atherosclerosis imaging. *Arterioscler Thromb Vasc Biol* 2019;**39**:841–849.
14. Senders ML, Calcagno C, Tawakol A, Nahrendorf M, Mulder WJM, Fayad ZA. PET/MR imaging of inflammation in atherosclerosis. *Nat Biomed Eng* 2023;**7**:202–220.
15. Garg PK, Bhatia HS, Allen TS, Grainger T, Pouncey AL, Dichek D, Virmani R, Golledge J, Allison MA, Powell JT. Assessment of subclinical atherosclerosis in asymptomatic people in vivo: measurements suitable for biomarker and Mendelian randomization studies. *Arterioscler Thromb Vasc Biol* 2024;**44**:24–47.
16. Mazhar F, Faucon AL, Fu EL, Szummer KE, Mathisen J, Gerward S, Reuter SB, Marx N, Mehran R, Carrero J-J. Systemic inflammation and health outcomes in patients receiving treatment for atherosclerotic cardiovascular disease. *Eur Heart J* 2024;**45**:4719–4730.
17. Farmakis D, Mueller C, Apple FS. High-sensitivity cardiac troponin assays for cardiovascular risk stratification in the general population. *Eur Heart J* 2020;**41**:4050–4056.
18. Williams SA, Kivimaki M, Langenberg C, Hingorani AD, Casas JP, Bouchard C, Jonasson C, Sarzynski MA, Shipley MJ, Alexander L, Ash J, Bauer T, Chadwick J, Datta G, DeLisle RK, Hagar Y, Hinterberg M, Ostroff R, Weiss S, Ganz P, Wareham NJ. Plasma protein patterns as comprehensive indicators of health. *Nat Med* 2019;**25**:1851–1857.
19. Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OT, Gunnarsdottir K, Olafsdottir TA, Halldorsson BV, Olason PI, Zink F, Gudjonsson SA, Sveinbjornsson G, Magnusson MI, Helgason A, Oddsson A, Halldorsson GH, Magnusson MK, Saevarsdottir S, Eiriksdottir T, Masson G, Stefansson H, Jonsdottir I, Holm H,

Rafnar T, Melsted P, Saemundsdottir J, Norddahl GL, Thorleifsson G, Ulfarsson MO, Gudbjartsson DF, Thorsteinsdottir U, Sulem P, Stefansson K. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* 2023;**622**:348–358.
20. Cui M, Cheng C, Zhang L. High-throughput proteomics: a methodological mini-review. *Lab Invest* 2022;**102**:1170–1181.
21. Gao S, Wang Z, Huang Y, Yang G, Wang Y, Yi Y, Zhou Q, Jian X, Zhao G, Li B, Xu L, Xia K, Tang B, Li J. Early detection of Parkinson's disease through multiplex blood and urine biomarkers prior to clinical diagnosis. *NPJ Parkinsons Dis* 2025;**11**:35.
22. Hallqvist J, Bartl M, Dakna M, Schade S, Garagnani P, Bacalini MG, Pirazzini C, Bhatia K, Schreglmann S, Xylaki M, Weber S, Ernst M, Muntean M-L, Sixel-Döring F, Franceschi C, Doykov I, Śpiewak J, Vinette H, Trenkwalder C, Heywood WE, Mills K, Mollenhauer B. Plasma proteomics identify biomarkers predicting Parkinson's disease up to 7 years before symptom onset. *Nat Commun* 2024;**15**:4759.
23. Guo Y, You J, Zhang Y, Liu WS, Huang YY, Zhang YR, Zhang W, Dong Q, Feng J-F, Cheng W, Yu J-T. Plasma proteomic profiles predict future dementia in healthy adults. *Nat Aging* 2024;**4**:247–260.
24. Dark HE, Paterson C, Daya GN, Peng Z, Duggan MR, Bilgel M, An Y, Moghekar A, Davatzikos C, Resnick SM, Loupy K, Simpson M, Candia J, Mosley T, Coresh J, Palta P, Ferrucci L, Shapiro A, Williams SA, Walker KA. Proteomic indicators of health predict Alzheimer's disease biomarker levels and dementia risk. *Ann Neurol* 2024;**95**:260–273.
25. Walker KA, Chen J, Shi L, Yang Y, Fornage M, Zhou L, Schlosser P, Surapaneni A, Grams ME, Duggan MR, Peng Z, Gomez GT, Tin A, Hoogeveen RC, Sullivan KJ, Ganz P, Lindbohm JV, Kivimaki M, Nevado-Holgado AJ, Buckley N, Gottesman RF, Mosley TH, Boerwinkle E, Ballantyne CM, Coresh J. Proteomics analysis of plasma from middle-aged adults identifies protein markers of dementia risk in later life. *Sci Transl Med* 2023;**15**:eadf5681.
26. Lyu J, Jiang M, Zhu Z, Wu H, Kang H, Hao X, Cheng S, Guo H, Shen X, Wu T, Chang J, Wang C. Identification of biomarkers and potential therapeutic targets for pancreatic cancer by proteomic analysis in two prospective cohorts. *Cell Genom* 2024;**4**:100561.
27. Papier K, Atkins JR, Tong TYN, Gaitskell K, Desai T, Ogamba CF, Parsaeian M, Reeves GK, Mills IG, Key TJ, Smith-Byrne K, Travis RC. Identifying proteomic risk factors for cancer using prospective and exome analyses of 1463 circulating proteins and risk of 19 cancers in the UK Biobank. *Nat Commun* 2024;**15**:4010.
28. Sun J, Liu Y, Zhao J, Lu B, Zhou S, Lu W, Wei J, Hu Y, Kong X, Gao J, Guan H, Gao J, Xiao Q, Li X. Plasma proteomic and polygenic profiling improve risk stratification and personalized screening for colorectal cancer. *Nat Commun* 2024;**15**:8873.
29. Broad-capture proteomics and machine learning for early detection of type 2 diabetes risk. *Nat Med* 2022;**28**:2261–2262.
30. Rooney MR, Chen J, Echouffo-Tcheugui JB, Walker KA, Schlosser P, Surapaneni A, Tang O, Chen J, Ballantyne CM, Boerwinkle E, Ndumele CE, Demmer RT, Pankow JS, Lutsey PL, Wagenknecht LE, Liang Y, Sim X, van Dam R, Tai ES, Grams ME, Selvin E, Coresh J. Proteomic predictors of incident diabetes: results from the Atherosclerosis Risk in Communities (ARIC) study. *Diabetes Care* 2023;**46**:733–741.
31. Luo H, Bauer A, Nano J, Petrera A, Rathmann W, Herder C, Hauck SM, Sun BB, Hoyer A, Peters A, Thorand B. Associations of plasma proteomics with type 2 diabetes and related traits: results from the longitudinal KORA S4/F4/FF4 study. *Diabetologia* 2023;**66**:1655–1668.
32. Jacobs BM, Vickaryous N, Giovannoni G, Proitsi P, Waters S, Dobson R. Plasma proteomic profiles of UK Biobank participants with multiple sclerosis. *Ann Clin Transl Neurol* 2024;**11**:698–709.
33. O'Neil LJ, Spicer V, Smolik I, Meng X, Goel RR, Anaparti V, Wilkins J, El-Gabalawy HS. Association of a serum protein signature with rheumatoid arthritis development. *Arthritis Rheumatol* 2021;**73**:78–88.
34. Argentieri MA, Xiao S, Bennett D, Winchester L, Nevado-Holgado AJ, Ghose U, Albukhari A, Yao P, Mazidi M, Lv J, Millwood I, Fry H, Rodosthenous RS, Partanen J, Zheng Z, Kurki M, Daly MJ, Palotie A, Adams CJ, Li L, Clarke R, Amin N, Chen Z, van Duijn CM. Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nat Med* 2024;**30**:2450–2460.
35. Helgason H, Eiriksdottir T, Ulfarsson MO, Choudhary A, Lund SH, Ivarsdottir EV, Hjorleifsson Eldjarn G, Einarsson G, Ferkingstad E, Moore KHS, Honarpour N, Liu T, Wang H, Hucko T, Sabatine MS, Morrow DA, Giugliano RP, Ostrowski SR, Pedersen OB, Bundgaard H, Erikstrup C, Arnar DO, Thorgeirsson G, Masson G, Magnusson OT, Saemundsdottir J, Gretarsdottir S, Steinthorsdottir V, Thorleifsson G, Helgadottir A, Sulem P, Thorsteinsdottir U, Holm H, Gudbjartsson D, Stefansson K. Evaluation of large-scale proteomics for prediction of cardiovascular events. *JAMA* 2023;**330**:725–735.
36. Ho FK, Mark PB, Lees JS, Pell JP, Strawbridge RJ, Kimenai DM, Mills NL, Woodward M, McMurray JJV, Sattar N, Welsh P. A proteomics-based approach for prediction of different cardiovascular diseases and dementia. *Circulation* 2025;**151**:277–287.
37. Kuku KO, Shearer JJ, Hashemian M, Oyetoro R, Park H, Dulek B, Bielinski SJ, Larson NB, Ganz P, Levy D, Psaty BM, Joo J, Roger VL. Development and validation of a protein risk score for mortality in heart failure: a community cohort study. *Ann Intern Med* 2024;**177**:39–49.
38. Royer P, Bjornson E, Adiels M, Josefson R, Hagberg E, Gummesson A, Bergström G. Large-scale plasma proteomics in the UK Biobank modestly improves prediction of major cardiovascular events in a population without previous cardiovascular disease. *Eur J Prev Cardiol* 2024;**31**:1681–1689.
39. Luo H, Huemer MT, Petrera A, Hauck SM, Rathmann W, Herder C, Koenig W, Hoyer A, Peters A, Thorand B. Association of plasma proteomics with incident coronary heart disease in individuals with and without type 2 diabetes: results from the population-based KORA study. *Cardiovasc Diabetol* 2024;**23**:53.
40. Park H, Norby FL, Kim D, Jang E, Yu HT, Kim TH, Uhm J-S, Sung J-H, Pak H-N, Lee M-H, Yang P-S, Joung B. Proteomic signatures for risk prediction of atrial fibrillation. *Circulation* 2025;**152**:217–229.
41. Xie R, Sha S, Brenner H, Schottker B. Circulating inflammation-related proteome improves cardiovascular risk prediction. Results from two large European cohort studies. *Eur J Epidemiol* 2025;**40**:1191–1203.
42. Nunez E, Fuster V, Gomez-Serrano M, Valdivielso JM, Fernandez-Alvira JM, Martinez-Lopez D, Rodríguez JM, Bonzon-Kulichenko E, Calvo E, Alfayate A, Bermudez-Lopez M, Escola-Gil JC, Fernández-Friera L, Cerro-Pardo I, Mendiguren JM, Sánchez-Cabo F, Sanz J, Ordovás JM, Blanco-Colio LM, García-Ruiz JM, Ibáñez B, Lara-Pezzi E, Fernández-Ortiz A, Martín-Ventura JL, Vázquez J. Unbiased plasma proteomics discovery of biomarkers for improved detection of subclinical atherosclerosis. *EBioMedicine* 2022;**76**:103874.
43. Baragetti A, Grigore L, Olmastroni E, Mattavelli E, Catapano AL. Plasma proteins associate with carotid plaques and predict incident atherosclerotic cardiovascular events. *Vascul Pharmacol* 2024;**156**:107394.
44. Baragetti A, Mattavelli E, Grigore L, Pellegatta F, Magni P, Catapano AL. Targeted plasma proteomics to predict the development of carotid plaques. *Stroke* 2022;**53**:e411–e414.
45. Royer P, Bjornson E, Adiels M, Alvez MB, Fagerberg L, Backhed F, Uhlén M, Gummesson A, Bergström G. Plasma proteomics for prediction of subclinical coronary artery calcifications in primary prevention. *Am Heart J* 2024;**271**:55–67.
46. Chen ML, Kho PF, Guarischi-Sousa R, Zhou J, Panyard DJ, Azizi Z, Gupte T, Watson K, Abbasi F, Assimes TL. Plasma proteomics and carotid intima-media thickness in the UK biobank cohort. *Front Cardiovasc Med* 2024;**11**:1478600.
47. Lind L, Gigante B, Borne Y, Feldreich T, Leppert J, Hedberg P, Östgren CJ, Nyström FH, Sundström J, Ärnlöv J, Baldassarre D, Tremoli E, Veglia F, Hamsten A, O'Donnell CJ, Franceschini N, Orho-Melander M, Nilsson J, Melander O, Engström G, Mälarstig A. Plasma protein profile of carotid artery atherosclerosis and atherosclerotic outcomes: meta-analyses and Mendelian randomization analyses. *Arterioscler Thromb Vasc Biol* 2021;**41**:1777–1788.
48. Chen QS, Bergman O, Ziegler L, Baldassarre D, Veglia F, Tremoli E, Strawbridge RJ, Gallo A, Pirro M, Smit AJ, Kurl S, Savonen K, Lind L, Eriksson P, Gigante B. A machine learning based approach to identify carotid subclinical atherosclerosis endotypes. *Cardiovasc Res* 2023;**119**:2594–2606.
49. Palmer LJ. UK Biobank: bank on it. *Lancet* 2007;**369**:1980–1982.
50. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;**12**:e1001779.
51. Liang J, Hu Z, Zhan C, Wang Q. Using propensity score matching to balance the baseline characteristics. *J Thorac Oncol* 2021;**16**:e45–e46.
52. Ho D, Imai K, King G, Stuart EA. Matchit: nonparametric preprocessing for parametric causal inference. *J Stat Softw* 2011;**42**:1–28.
53. Coffey S, Lewandowski AJ, Garratt S, Meijer R, Lynum S, Bedi R, Paterson J, Yaqub M, Noble JA, Neubauer S, Petersen SE, Allen N, Sudlow C, Collins R, Matthews PM, Leeson P. Protocol and quality assurance for carotid imaging in 100,000 participants of UK Biobank: development and assessment. *Eur J Prev Cardiol* 2017;**24**:1799–1806.
54. Holle R, Happich M, Lowel H, Wichmann HE, Group MKS. KORA–a research platform for population based health research. *Gesundheitswesen* 2005;**67**:S19–S25.
55. Wik L, Nordberg N, Broberg J, Bjorkesten J, Assarsson E, Henriksson S, Grundberg I, Pettersson E, Westerberg C, Liljeroth E, Falck A, Lundberg M. Proximity extension assay in combination with next-generation sequencing for high-throughput proteome-wide analysis. *Mol Cell Proteomics* 2021;**20**:100168.
56. Dhindsa RS, Burren OS, Sun BB, Prins BP, Matelska D, Wheeler E, Mitchell J, Oerton E, Hristova VA, Smith KR, Carss K, Wasilewski S, Harper AR, Paul DS, Fabre MA, Runz H, Viollet C, Challis B, Platt A, Vitsios D, Ashley EA, Whelan CD, Pangalos MN, Wang Q, Petrovski S. Rare variant associations with plasma protein levels in the UK Biobank. *Nature* 2023;**622**:339–347.
57. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, Hou L, Kvikstad EM, Burren OS, Davitte J, Ferber KL, Gillies CE, Hedman ÅK, Hu S, Lin T, Mikkilineni R, Pendergrass RK, Pickering C, Prins B, Baird D, Chen C-Y, Ward LD, Deaton AM, Welsh S, Willis CM, Lehner N, Arnold M, Wörheide MA, Suhre K, Kastenmüller G, Sethi A, Cule M, Raj A, Burkitt-Gray L, Melamud E, Black MH, Fauman EB, Howson JMM, Kang HM, McCarthy MI, Nioi P, Petrovski S, Scott RA, Smith EN,

Szalma S, Waterworth DM, Mitnaul LJ, Szustakowski JD, Gibson BW, Miller MR, Whelan CD. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* 2023;**622**:329–338.
58. Niu L, Thiele M, Geyer PE, Rasmussen DN, Webel HE, Santos A, Gupta R, Meier F, Strauss M, Kjaergaard M, Lindvig K, Jacobsen S, Rasmussen S, Hansen T, Krag A, Mann M. Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat Med* 2022;**28**:1277–1287.
59. Omarov M, Zhang L, Jorshery SD, Malik R, Das B, Bellomo TR, Mansmann U, Menten MJ, Natarajan P, Dichgans M, Kalic M, Raghu VK, Berger K, Anderson CD, Georgakis MK. Deep learning-based detection of carotid plaques informs cardiovascular risk prediction and reveals genetic drivers of atherosclerosis. medRxiv 2024.10.17.24315675.
60. Touboul PJ, Hennerici MG, Meairs S, Adams H, Amarenco P, Bornstein N, Csiba L, Desvarieux M, Ebrahim S, Hernandez Hernandez R, Jaff M, Kownator S, Naqvi T, Prati P, Rundek T, Sitzer M, Schminke U, Tardif J-C, Taylor A, Vicaut E, Woo KS. Mannheim carotid intima-media thickness and plaque consensus (2004-2006-2011). An update on behalf of the advisory board of the 3rd, 4th and 5th watching the risk symposia, at the 13th, 15th and 20th European Stroke Conferences, Mannheim, Germany, 2004, Brussels, Belgium, 2006, and Hamburg, Germany, 2011. *Cerebrovasc Dis* 2012;**34**:290–296.
61. Hundemer GL, White CA, Norman PA, Knoll GA, Tangri N, Sood MM, Hiremath S, Burns KD, McCudden C, Akbari A. Performance of the 2021 race-free CKD-EPI creatinine- and cystatin C-based estimated GFR equations among kidney transplant recipients. *Am J Kidney Dis* 2022;**80**:462–472.e461.
62. Peters A, Doring A, Ladwig KH, Meisinger C, Linkohr B, Autenrieth C, Baumeister S E, Behr J, Bergner A, Bickel H, Bidlingmaier M, Dias A, Emeny R T, Fischer B, Grill E, Gorzelniak L, Hänsch H, Heidbreder S, Heier M, Horsch A, Huber D, Huber R M, Jörres R A, Kääb S, Karrasch S, Kirchberger I, Klug G, Kranz B, Kuch B, Lacruz M E, Lang O, Mielck A, Nowak D, Perz S, Schneider A, Schulz H, Müller M, Seidl H, Strobl R, Thorand B, Wende R, Weidenhammer W, Zimmermann A-K, Wichmann H-E, Holle R. [Multimorbidity and successful aging: the population-based KORA-Age study]. *Z Gerontol Geriatr* 2011;**44**:41–54.
63. Meisinger C, Thorand B, Schneider A, Stieber J, Doring A, Lowel H. Sex differences in risk factors for incident type 2 diabetes mellitus: the MONICA Augsburg cohort study. *Arch Intern Med* 2002;**162**:82–89.
64. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;**42**:2439–2454.
65. Ritchie SC, Jiang X, Pennells L, Xu Y, Coffey C, Liu Y, Gibson JT, Surendran P, Karthikeyan S, Lambert SA, Danesh J, Butterworth AS, Wood A, Kaptoge S, Di Angelantonio E, Inouye M. Combined clinical, metabolomic, and polygenic scores for cardiovascular risk prediction. *Eur Heart J* 2026;**47**:1861–1873.
66. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, Weeks EM, Wang M, Hindy G, Zhou W, Grace C, Roselli C, Marston NA, Kamanu FK, Surakka I, Venegas LM, Sherliker P, Koyama S, Ishigaki K, Åsvold BO, Brown MR, Brumpton B, de Vries PS, Giannakopoulou O, Giardoglou P, Gudbjartsson DF, Güldener U, Haider SMI, Helgadottir A, Ibrahim M, Kastrati A, Kessler T, Kyriakou T, Konopka T, Li L, Ma L, Meitinger T, Mucha S, Munz M, Murgia F, Nielsen JB, Nöthen MM, Pang S, Reinberger T, Schnitzler G, Smedley D, Thorleifsson G, von Scheidt M, Ulirsch JC, Arnar DO, Burtt NP, Costanzo MC, Flannick J, Ito K, Jang D-K, Kamatani Y, Khera AV, Komuro I, Kullo IJ, Lotta LA, Nelson CP, Roberts R, Thorgeirsson G, Thorsteinsdottir U, Webb TR, Baras A, Björkegren JLM, Boerwinkle E, Dedoussis G, Holm H, Hveem K, Melander O, Morrison AC, Orho-Melander M, Rallidis LS, Ruusalepp A, Sabatine MS, Stefansson K, Zalloua P, Ellinor PT, Farrall M, Danesh J, Ruff CT, Finucane HK, Hopewell JC, Clarke R, Gupta RM, Erdmann J, Samani NJ, Schunkert H, Watkins H, Willer CJ, Deloukas P, Kathiresan S, Butterworth AS. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet* 2022;**54**:1803–1815.
67. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang S-J, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen L-P, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han B-G, Huang J, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki M-L, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon F-U-R, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim B-J, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJF, Melander O, Metspalu A, März W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;**47**:1121–1130.
68. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;**44**:W90–W97.
69. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;**369**:1318–1330.
70. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, AK: Association for Computing Machinery, 2019. pp. 2623–2631.
71. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA: Curran Associates Inc., 2017. pp. 4768–4777.
72. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. Catboost: unbiased boosting with categorical features. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R (eds.), *Advances in Neural Information Processing Systems 31*. Red Hook, NY: Curran Associates, Inc.; 2018. pp. 6638–6648.
73. Inoue E. nricens: NRI for Risk Prediction Models with Time to Event and Binary Response Data. R package version 1.6; 2018. https://CRAN.R-project.org/package=nricens (1 February 2026, date last accessed).
74. Uno H, Tian L, Cai T, Kohane IS, Wei L. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med* 2013;**32**:2430–2442.
75. Qiu W, Chavarro J, Lazarus R, Rosner B, Ma J. powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies. R package version 0.1.3; 2021. https://CRAN.R-project.org/package=powerSurvEpi (1 February 2026, date last accessed).
76. Bell KJL, White S, Hassan O, Zhu L, Scott AM, Clark J, Glasziou P. Evaluation of the incremental value of a coronary artery calcium score beyond traditional cardiovascular risk assessment: a systematic review and meta-analysis. *JAMA Intern Med* 2022;**182**:634–642.
77. Peterson TE, Lima JAC, Shah SJ, Bluemke DA, Bertoni AG, Liu Y, Ngo D, Varadarajan V, Mychaleckyj JC, Johnson CW, Psaty BM, Clish CB, Taylor KD, Durda P, Tracy RP, Gerszten RE, Rich SS, Rotter JI, Post WS, Pankow JS. Proteomics of left ventricular structure in the Multi-Ethnic Study of Atherosclerosis. *ESC Heart Fail* 2025;**12**:239–249.
78. Zusi C, Csermely A, Rinaldi E, Bertoldo K, Bonetti S, Boselli ML, Travia D, Bonora E, Bonadonna RC, Trombetta M. Crosstalk between genetic variability of adiponectin and leptin, glucose-insulin system and subclinical atherosclerosis in patients with newly diagnosed type 2 diabetes. The Verona Newly Diagnosed Type 2 Diabetes Study 14. *Diabetes Obes Metab* 2023;**25**:2650–2658.
79. Ciccone M, Vettor R, Pannacciulli N, Minenna A, Bellacicco M, Rizzon P, Giorgino R, De Pergola G. Plasma leptin is independently associated with the intima-media thickness of the common carotid artery. *Int J Obes Relat Metab Disord* 2001;**25**:805–810.
80. Hobaus C, Pesau G, Herz CT, Wrba T, Koppensteiner R, Schernthaner GH. Angiopoietin-2 and survival in peripheral artery disease patients. *Thromb Haemost* 2018;**118**:791–797.
81. Gorla G, Potenza A, Carrozzini T, Pollaci G, Acerbi F, Vetrano IG, Ferroli P, Canavero I, Rifino N, Bersano A, Gatti L. Angiopoietin-2 associates with poor prognosis in Moyamoya angiopathy. *Ann Clin Transl Neurol* 2024;**11**:1590–1603.
82. Duan S, Luo X, Dong C. Identification of susceptibility modules for coronary artery disease using a genome wide integrated network analysis. *Gene* 2013;**531**:347–354.
83. Talior-Volodarsky I, Connelly KA, Arora PD, Gullberg D, McCulloch CA. Alpha11 integrin stimulates myofibroblast differentiation in diabetic cardiomyopathy. *Cardiovasc Res* 2012;**96**:265–275.
84. Mohanraj L, Kim HS, Li W, Cai Q, Kim KE, Shin HJ, Lee Y-J, Lee WJ, Kim JH, Oh Y. IGFBP-3 inhibits cytokine-induced insulin resistance and early manifestations of atherosclerosis. *PLoS One* 2013;**8**:e55084.
85. Manta CP, Leibing T, Friedrich M, Nolte H, Adrian M, Schledzewski K, Krzistetzko J, Kirkamm C, David Schmid C, Xi Y, Stojanovic A, Tonack S, de la Torre C, Hammad S, Offermanns S, Krüger M, Cerwenka A, Platten M, Goerdt S, Géraud C. Targeting of scavenger receptors stabilin-1 and stabilin-2 ameliorates atherosclerosis by a plasma proteome switch mediating monocyte/macrophage suppression. *Circulation* 2022;**146**:1783–1799.
86. Gorabi AM, Kiaie N, Khosrojerdi A, Jamialahmadi T, Al-Rasadi K, Johnston TP, Sahebkar A. Implications for the role of lipopolysaccharide in the development of atherosclerosis. *Trends Cardiovasc Med* 2022;**32**:525–533.
87. Sallam T, Ito A, Rong X, Kim J, van Stijn C, Chamberlain BT, Jung ME, Chao LC, Jones M, Gilliland T, Wu XH, Su GL, Tangirala RK, Tontonoz P, Hong C. The macrophage LBP gene is an LXR target that promotes macrophage survival and atherosclerosis. *J Lipid Res* 2014;**55**:1120–1130.

88. Singh B, Tzoulaki I, Mayr M. Precision medicine requires precision proteomics: discordance between proteomic and clinical assays in UK Biobank. *Cardiovasc Res* 2025;**121**:2293–2295.
89. Maretty L, Gill D, Simonsen L, Soh K, Zagkos L, Galanakis M, Sibbesen J, Iglesias MT, Secher A, Valkenborg D, Purnell JQ, Knudsen LB, Tahrani AA, Geybels M. Proteomic changes upon treatment with semaglutide in individuals with obesity. *Nat Med* 2025;**31**:267–277.
90. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med* 2022;**28**:583–590.
91. Katsoularis I, Fonseca-Rodriguez O, Farrington P, Lindmark K, Fors Connolly AM. Risk of acute myocardial infarction and ischaemic stroke following COVID-19 in Sweden: a self-controlled case series and matched cohort study. *Lancet* 2021; **398**:599–607.
92. Arrigo M, Vodovar N, Von Moos S, Masson E, Segerer S, Cippa PE, Mebazaa A. High accuracy of proximity extension assay technology for the quantification of plasma brain natriuretic peptide. *J Clin Lab Anal* 2018;**32**:e22574.
93. Skau E, Wagner P, Leppert J, Arnlov J, Hedberg P. Are the results from a multiplex proteomic assay and a conventional immunoassay for NT-proBNP and GDF-15 comparable? *Clin Proteomics* 2023;**20**:5.
94. Blum MF, Chen J, Surapaneni A, Turner ST, Ballantyne CM, Welling PA, Köttgen A, Coresh J, Crews DC, Grams ME. Renin: measurements, correlates, and associations with long-term adverse kidney outcomes. *Am J Hypertens* 2023;**36**:42–49.